INVESTIGATION OF THE MUTATIONAL CONSEQUENCES OF DEFECTIVE MISMATCH REPAIR IN HUMANS

Andrea Louise Bacon

Thesis submitted for the degree

of Doctor in Philosophy

The University of Edinburgh

2001

Dedication

I dedicate this thesis to my Mum and Dad for all their love and support and also to my sister Karolyn.

Declaration

I declare that this thesis was composed entirely by myself and that the research presented is my own unless otherwise stated.

Andrea Louise Bacon

September 2001

The following publications derived from the research presented in this thesis:

Bacon, A.L., Farrington, S.M., and Dunlop, M.G. (2000). Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells. Hum. Mol. Genet., *9*: 2707-2713.

Bacon, A.L., Dunlop, M.G., and Farrington, S.M. (2001) Hypermutability of a poly(A/T) tract in the human germline. Nucleic Acids Res., *29*:4405-4413

Bacon, A.L., Farrington, S.M., and Dunlop, M.G. (2001) Mutation frequency in coding and non-coding repeat sequences in mismatch repair deficient cells derived from normal human tissue. Oncogene, *20*:7464-7471

Abstract

Bacon, A.L., Farrington, S.M. and Dunlop, M.G. (2000) Genetic instability in normal cells resulting in mutation of both microsatellite and coding sequences. Proc. Amer. Assoc. Canc. Res., *41*:A2244.

Abstract

Mismatch repair (MMR) defects results in widespread instability at repetitive DNA in both coding and non-coding sequences. Repetitive tracts within the coding sequences of genes including *TGFBR2* and *BAX* are frequently mutated in MMR deficient cancers and implicated in tumour progression. The Wnt pathway genes *APC* and β-catenin, are commonly mutated in MMR proficient cancers and are also mutated in a lower proportion of MMR deficient cancers, suggesting they may also be prone to mutation as a consequence of MMR defects. However, there has been little study of the relative contribution of inherent instability and selection pressure at such sequences in neoplasms. One explanation as to why sequences are mutated frequently in MMR deficient colorectal cancer (CRC) is that they are inherently prone to mutations. To investigate this hypothesis, MMR deficient cells (lbl-1260, lbl-1261) derived from normal tissue (B-lymphocytes), have been utilised. Analysing these cells minimises the effects of selection pressures that bias the apparent frequency of mutations in cancer cells and "unmasks" inherent stability of given sequences. Using this system, the research aims were to identify whether factors independent of tumourigenesis, influence the manifestation of the mutator phenotype at repetitive tracts. Also, whether mutations in the coding sequences of genes implicated in tumourigenesis, can occur as a consequence of MMR defects when selection pressure is minimal.

A SP-PCR technique was employed and demonstrated that lbl-1261 and lbl-1260 exhibit microsatellite instability. At the D2S123 locus, the sequence of the constitutional allele itself is shown to be a determinant of instability in the absence of MMR activity. Presence of an interruption within the (CA)n repeat, is common within the Scottish population and confers differential stability at individual microsatellite alleles in MMR deficient cells. At a second microsatellite locus (BAT-40), extreme sensitivity to mutation in the presence of MMR deficiency is observed. This is further indicated by analysis of the mutation frequency within families and also from sperm DNA where germline instability at this poly(A) locus is revealed. A combination of SP-PCR, allele cloning and restriction digest methods were employed to determine whether differences in mutation frequencies of the *TGFBR2* and *BAX*

genes in MMR deficient CRC, may be due to differential stability in the presence of MMR defects. MMR deficiency is shown to be associated with an excess of mutations at the poly(A)$_{10}$ tract of *TGFBR2* in cells of a normal lineage. Surprisingly, at the poly(G)$_8$ tract of the *BAX* gene, an excess of mutations is not observed. The inherent stability of Wnt pathway gene sequences frequently mutated in MMR proficient CRC was addressed. Mutations of *CTNNB1* and *APC* are observed in a small proportion of MMR deficient CRCs. However, despite rigorous mutation analysis of exon 3 of the *CTNNB1* gene, there is no indication that this region is particularly prone to mutation consequent of MMR defects. Extensive analysis of inherent mutability was also performed for the APC gene. Interestingly, an unexpectedly high level of instability was detected in exon 15 in both normal and cancer cells, proficient in MMR. But in lbl-1261 the instability was markedly lower suggestive of complex relationship between MMR defects and mutation frequency at this locus.

In this thesis, the importance of inherent instability in the accumulation of mutational events observed in MMR deficient CRC has been investigated. The work shows conclusively that the nature of the repeat in non-coding and coding sequences affects the manifestation of the mutator phenotype. Differences in the mutation frequency of genes mutated in CRC are influenced by inherent instability at such sequences in the presence of MMR defects even in non-neoplastic cells.

Acknowledgements

I wish to thank my supervisor Prof. Malcolm Dunlop for his support during the course of this research. Particular thanks to our great leader for keeping me focused, for critically reading my manuscripts and for letting me spend lots of money! A big round of applause must also go to Dr. Susan Farrington for her continued help, advice and encouragement throughout my PhD. Also for feeding me too many treats and for buying champagne! Thanks very much.

Id also like to thank Sheila Mcbeath for her help in maintaining the cell lines used in this study and for passing on a few tricks that prevented me killing all the cells I tried to grow!

Many thanks to everyone else in the lab that have assisted with various aspects of this work, and who have endured my pestering without complaint. A special gold star must go to all the Dunlop girls, past and present who have made my time in the lab a pleasure, have kept me smiling, provided entertainment, and generally taken good care of their PhD student. Cheers for helping me retain some sanity, especially during the last few months!

This PhD also owes its success to all those that have made my time in Edinburgh lots of fun. Thanks to Nic, not only for being a great flatmate and friend but for also being superb in co-organising the fine catalogue of parties that have spanned the course of this PhD. The list of friends who have participated in the dressing up and the dancing, over the last four years is long. You know who you are and id like to take this opportunity to thank you most warmly for your valued contribution to this thesis. Thanks also to the continued friendship from the Essex and Nottingham girls.

An extra special thank you must go to Mum, Dad, Kazz and Nan for their ongoing love, encouragement and support. Your interest and enthusiasm in all that I do means a great deal, is very much appreciated and undoubtedly keeps me going.

Finally a massive thank you to Martin not only for advice and assistance with my work but also for being pleased for me when things have gone well, for giving me hugs when things have not and generally for being brilliant. I could not have done it without you.

Abbreviations

A	Adenosine
aa	Amino acid
ACF	Aberrant crypt foci
ACN	Acetonitrile
ADP	Adenine di phosphate
APC	Adenomatous polyposis coli
APS	Ammonium persulphate
ASR	Age standardised rates
ATP	Adenine tri phosphate
bp	Base pairs of DNA
BSA	Bovine serum albumin
C	Cytosine
CEPH	Centre D'Etude Du Polymorphisme Humain
CFTR	Cystic fibrosis transmembrane conductance regulator
CRC	Colorectal cancer
CTNNB1	Gene encoding for β-catenin
dH_2O	Distilled water
DHPLC	Denaturing high performance liquid chromotography
DNA	Deoxyribonucleic acid
dNTP	Deoxy nucleotide tri-phosphate
EBV	Epstein barr virus
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetra-acetic acid

FACS	Fluorescence assisted cell sorting
FAP	Familial adenomatous polyposis coli
FAM	6-carboxyfluorescein
FCS	Fetal Calf serum
FMR1	Fragile X gene
G	Guanine
HEPES	N-2 hydroxyethylpiperazine-N'-2-ethanesulfonic acid
HEX	4, 7, 2', 4', 5', 7'-hexachloro-6-carboxyfluorescein
HNPCC	Hereditary non-polyposis colorectal cancer
3-β-HSD	3-beta-hydroxysteroid dehydrogenase
IPTG	β-D-thioglactopyranoside
Kb	Kilobase
Lbl	Lymphoblast cell line
LOH	Loss of heterozygosity
MCR	Mutation cluster region
MD-	Laboratory sample identification
MgCl	Magnesium chloride
MLH1	Homologue of the bacterial MutL mismatch repair protein
MLH3	Homologue of the bacterial MutL mismatch repair protein
MMR	Mismatch repair
MNNG	N-methyl-N'-nitro-N-nitrosanguanidine
MNU	N-methyl-N-nitosourea
mRNA	Messenger ribonucleic acid
MSH2	Homologue of bacterial MutS mismatch repair protein
MSH3	Homologue of bacterial MutS mismatch repair protein

MSH6	Homologue of bacterial MutS mismatch repair protein
MSI	Microsatellite instability
MSI^+	Microsatellite unstable
MSI-H	High-level microsatellite instability
MSI-L	Low-level microsatellite instability
MSS	Microsatellite stable
MTS	Muir-Torre syndrome
NaCl	Sodium chloride
NaOAc	Sodium Acetate
NSAID	Non-steroidal anti-inflammatory drug
Nt	Nucleotide
O^6-MeG	O^6-methyl-guanine
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PMS1	Homologue of bacterial MutL mismatch repair protein
PMS2	Homologue of bacterial MutL mismatch repair protein
PJS	Peutz-Jehger syndrome
rpm	Revolutions per minute
RNA	Ribonucleic acid
RT-PCR	Reverse transcription and PCR amplification
SCA1	Spinocerebella atxia type 1
S.cerevisiae	*Saccharomyces cerevisiae*
SDS	Sodium dodecyl sulphate
SP-PCR	Small pool polymerase chain reaction

SSC	Salt and sodium citrate buffer
SSR	Simple sequence repeat
T	Thymine
TAE	Tris Acetate EDTA buffer
TEAA	Trimethyl ammonium acetate
TEMED	N, N, N', N' tetramethyl-1-2-diaminomethane
TGFBR2	Transforming growth factor beta type 2 receptor
Tris	Tris(hydroxymethyl)aminomethane
TS	Turcots syndrome
TV	Trypsin versene
UTR	Untranslated region
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactosidase

Contents

Chapter 1 Introduction

Chapter 2 Materials and Methods

Chapter 3 Characterisation of Cell Lines Derived from Normal Tissue with Defects in MMR

Chapter 4 MMR Defects in Cells Derived from Normal Tissue Are Associated with Microsatellite Instability

Chapter 6 Hypermutability at the BAT-40 poly(A/T) Tract in the Human Germline

Chapter 7 Susceptibility of *TGBFR2* and *BAX* Coding Repeat Sequences to Mutation in Non-Cancer Derived Cells with MMR Defects

Chapter 9 Discussion and Summary

Chapter 10 Bibliography

Appendix A

Published Articles

List of Tables and Figures

Chapter 1

Introduction

1.1 Introduction

The research described in this thesis has addressed how defects in mismatch repair (MMR) and the consequent elevation in mutation rate, contributes to the initiation and progression of colorectal cancer (CRC). In particular this project has investigated how inherent instability at both non-coding and coding sequence can influence the frequency with which mutations accumulate in cancers with MMR defects. Specific factors independent of the tumourigenic process itself, which affect the propensity to mutation, have also been determined.

Inherent instability and selection pressures are believed to be important influences in tumourigenesis. However, the specific consequences of these factors and the relative contribution made to the acquisition of genetic changes observed in cancers, are not well understood. Therefore, the examination of molecular events that arise exclusively as a result of MMR defects has considerable relevance to understanding the effects of inherent instability and increased mutation rate in cancer development.

The significance in elucidating the molecular mechanisms that underlie CRC is underscored by the fact that cancer of the colon is a major health problem. In this chapter the incidence of CRC is discussed and the contribution from aetiological and genetic factors is reviewed. Evidence that an increase in instability is an important and necessary feature of cancer is presented. In addition, the notion that selection pressure represents the major driving force in tumourigenesis is also considered. Malignancies associated with MMR defects are described, to exemplify the spectrum of cancers in which genomic instability is involved. This includes a description of syndromes associated with extra-colonic cancers. The mechanism by which MMR maintains genome integrity is detailed, with emphasise on how inactivation of this system results in the drastic elevation of mutations that are observed in the associated cancers. The genetic and epigenetic mechanisms of inactivation of MMR are then

presented. Current understanding of the consequences of MMR defects in terms of widespread microsatellite instability and the manifestation of a mutator phenotype is described and the evidence suggesting that instability results in tumourigenesis, via mutation of specific cancer causing genes is detailed. Finally the aims of the project are presented and the experimental approach is described.

1.2 Environmental and Genetic Factors in Colorectal Cancer

1.2.1 Incidence of colorectal cancer

Cancers of the colon and rectum (CRCs) represent a major health problem in the western world. Despite advances in modern medicine the prevalence of CRC remains second only to lung cancer in developed countries (Parkin *et al.*, 1999). In the United Kingdom (UK) there are 34,000 new cases of CRC and more than 20,000 deaths annually (Farrington and Dunlop, in press) presenting a significant financial burden to the National Health Service. In the United States (US), CRC accounts for 10% of all cancer mortality (Pisani *et al.*, 1999). In 2001 alone there will be an estimated 135,400 new cases in the US including, 98,200 of colon cancer and 37,200 of rectal cancer with an estimated 56,700 deaths (American Cancer Society, www.cancer.org). The prognosis for patients with CRC is relatively poor. Data from the Scottish information and statistics division reveal that an average of just 44.5% of patients are still alive 5 years after presentation (ISD Scotland; www.show.scot.nhs.uk/isd/index.htm). CRC is the only cancer to occur with approximately equal frequencies in men and women (Parkin *et al.*, 1999). Although, the age adjusted rates in men exceed woman by as much as 20%, in high incidence areas such as North America and Australia (Parkin *et al.*, 1999).

Colorectal cancer has a multifactorial aetiology that involves environmental factors and genetic susceptibility. Environmental factors are evidenced by the marked differences in world prevalence of the disease. Carcinoma of the colon and rectum is predominantly a disease of westernised societies (Burkitt, 1971; Haenszel and Correa, 1971; Parkin *et al.*, 1999; Pisani *et al.*, 1999) and age standardised rates (ASR) suggest that in industrialised countries, bowel cancer is more than fifteen

times more common than it is in developing countries (Burkitt, 1971; Parkin *et al.*, 1999). Australia and New Zealand have the highest ASR of CRC incidence (89/100,000) (Parkin *et al.*, 1999), whereas the lowest rates are reported in central Africa (5/100,000) (Haenszel and Correa, 1971; Parkin *et al.*, 1999; Potter, 1999).

CRC has long been known to occur more frequently in certain families and some 20-25% of all cases are hereditary (Bonelli *et al.*, 1988). A number of hereditary syndromes have now been defined and the last ten years has seen an increase in the number of cancer predisposition genes being isolated, which have moderate to high risk penetrance (Farrington and Dunlop, in press).

1.2.2 Aetiological factors in colorectal cancer

The increased incidence of CRC in developed countries indicates a contribution from environmental factors. This is exemplified by studies of migrants from low to high incidence countries, who attain cancer incidence rates similar to those of their adopted country within a single generation (Haenszel, 1961; Stemmermann, 1970 McMichael and Giles, 1988). American Africans now have an incidence of bowel cancer that is comparable to that of Caucasians rather than that of native Africans (Burkitt, 1971; Parkin *et al.*, 1999). In addition, Japanese who have immigrated to America have a raised incidence of bowel cancer, compared to that observed in Japan (Burkitt, 1971; Stemmermann, 1970). In most cases, these migrants have accepted the dietary customs of their country of adoption indicating that the 15-fold international difference may be explained in large part by dietary factors. Furthermore rates of CRC incidence are increasing in countries with previously low rates, possibly due to lifestyle changes (Parkin *et al.*, 1999; Potter, 1999).

Although there is little doubt that diet and/or lifestyle are importance risk factors contributing to the high incidence of CRC in the western world, there is both contradictory and negative evidence regarding what the exact risks may be. Numerous studies have addressed the causality of various dietary and lifestyle components ranging from carcinogenic effects of smoking, alcohol and excess meat and fats as well as protective effects from fruit and fibre intake and exercise. These

studies have been reviewed extensively elsewhere (Gertig and Hunter, 1998; Potter, 1999) but the main themes will be presented here.

One of the most notable differences between the diet of the western world where bowel cancer is most prevalent, and that of less developed countries, is the proportions of un-absorbable fibre and refined carbohydrate in the food ingested. Burkitt, originally proposed that a decrease in dietary fibre in western diets might act by increasing bowel transit time and increasing the faecal concentrations of carcinogens (Burkitt, 1971). However, while the consumption of fibre in terms of increased fruit and vegetable intake has been related to lower risk of colon cancer (Willett, 1989; Trock *et al.*, 1990; Levi *et al.*, 1999), the consumption of cereal products has not (Willett, 1989). In addition, Giovannucci *et al* reported no inverse relationship with either vegetables or fibre and CRC incidence in a prospective study of over 50,000 US men (Giovannucci *et al.*, 1994), indicating that while there is strong evidence that fruit and vegetables are protective this is not necessarily conclusive.

Increased consumption of red meat has also been associated with increased risk of CRC in prospective and case-control studies (Willett, 1989; Willett *et al.*, 1990 Giovannucci *et al.*, 1994; Levi *et al.*, 1999). The association has been explained in terms of its fat content, bile acid production or by carcinogens developed by cooking (Willett, 1989; Willett *et al.*, 1990). However, there are also negative studies and a number of European reports have shown no consistent association between red meat consumption and CRC (Goldbohm *et al.*, 1994; Franceschi *et al.*, 1997).

Red meat and vegetable intake represent the two best-recognised dietary correlates of CRC (Willett, 1989; Potter, 1999). Nonetheless, the possibility of an association of CRC with other dietary and lifestyle factors such as alcohol consumption, physical activity, smoking and NSAID intake has been suggested.

Levi *et al* found a moderate association between alcohol drinking and CRC in a case-control study comprising 223 patients and this supported epidemiological evidence from Longnecker *et al* who concluded a weak overall association (Longnecker *et al.*, 1990; Levi *et al.*, 1999). Le Marchand *et al* evaluated the contribution of a number of factors predominantly associated with a western lifestyle

to CRC, including that from increased alcohol intake (Le Marchand *et al.*, 1997). A correlation between alcohol and CRC incidence in Asian immigrants to the US was observed (Le Marchand *et al.*, 1997). Le Marchand *et al* also reported that obesity, high caloric intake and little physical activity were each independently associated with an increased risk of CRC (Le Marchand *et al.*, 1997).

Long-term aspirin intake has been consistently demonstrated to have a preventative role in cancer risk (Muscat *et al.*, 1994; La Vecchia *et al.*, 1997). The mechanism for this protective effect has not been precisely established although evidence suggests that the anti neoplastic activity of aspirin and other NSAIDS is predominantly mediated by inhibition of the cyclooxygenase (COX) enzymes (Williams *et al.*, 1999). However, it has also been suggested that aspirin induces the activation of NF-KB, necessary for its anti tumour activity and this may also contribute to the protective effect of aspirin that has been observed (Stark *et al.*, 2001).

In spite of the evidence that non-genetic influences account for high rates of colon cancer in industrialised countries, few factors have been definitively proven to offer the potential for prevention. Current evidence suggests that a variety of exogenous agents (e.g., alcohol intake and meat consumption) may increase the risk while others (e.g. NSAIDS and vegetables) may reduce the risk (Potter, 1999). However, it is clear that the potential for preventing CRC by changes in diet needs to be more clearly defined.

1.2.3 Genetics factors in colorectal cancer

The multifactorial aetiology of CRC involves genetic susceptibility as well as environmental factors. Up to one quarter of all CRC cases are associated with a family history (Bonelli *et al.*, 1988). In addition, common genetic changes in subsets of sporadic CRCs have also been elucidated (Fearon and Vogelstein, 1990). The last ten years has seen significant advances in understanding the key molecular events involved in susceptibility to CRC and these are now beginning to have real clinical impact on the disease.

Hereditary non-polyposis colorectal cancer (HNPCC) represents one of the major familial CRC syndromes. HNPCC accounts for 5-8% of cancers of the colon (Mecklin *et al.*, 1995; Lynch and de la Chapelle, 1999). However, other reports on the percentage of CRC caused by HNPCC have provided varying figures of between ~0.5-10% (Aaltonen *et al.*, 1994; Bellacosa *et al.*, 1996; Aaltonen *et al.*, 1998; Cunningham *et al.*, 2001). Highly penetrant germline mutations in MMR genes are associated with tumour microsatellite instability (MSI) and are responsible for the majority of HNPCC cases (Aaltonen *et al.*, 1993; Aaltonen *et al.*, 1998). The clinical aspects and genetics of this disorder are discussed in detail later.

Germline mutations in the adenomatous polyposis coli (*APC*) gene also confer an extremely high risk of developing CRC as part of the disorder, familial adenomatous polyposis (FAP) (Groden *et al.*, 1991; Joslyn *et al.*, 1991; Kinzler *et al.*, 1991; Nishisho *et al.*, 1991). In contrast to HNPCC, this disease is characterised by hundreds to thousands of colorectal polyps that develop at an early age. Although these polyps are benign, their frequency and early onset invariably means that one or more will progress to malignancy (Kinzler and Vogelstein, 1996). However, germline mutations at the *APC* gene are somewhat less frequent than those in MMR genes and FAP accounts for less than 1% of CRC (Bisgaard *et al.*, 1994; Farrington and Dunlop, in press).

Other gene mutations have been elucidated that underlie a number of rarer disorders, associated with an elevated risk of CRC although the risk of CRC is markedly lower in these syndromes compared to either HNPCC or FAP. The clinical phenotypes of these disorders are heterogeneous, but each is characterised by benign polyps of the intestine. The rare autosomal dominant disorder Puetz-Jeghers syndrome (PJS) is associated with mutations of a novel serine threoinine kinase known as *LKB1* or *STK1* (Hemminki *et al.*, 1998; Jenne *et al.*, 1998). Mutations of the putative tyrosine phosphotase gene *PTEN* have been identified in families with the autosomal dominant disorder, Cowdens disease (Liaw *et al.*, 1997) and also within a small number of Juvenile polyposis coli families (Lynch *et al.*, 1997; Olschwang *et al.*, 1998). Juvenile polyposis coli, which has a very early age of onset, usually under 10 years, has also been associated with germline mutations of *SMAD4/DPC4* in a number of families (Howe *et al.*, 1998).

In addition to genes predisposing to an increased risk of CRC, somatic genetic alterations arising during colorectal tumour development have been identified and it is established that cancer cells contain many genetic abrogation's (Peinado *et al.*, 1992; Lengauer *et al.*, 1998). The sequence from adenoma to carcinoma is well understood on a clinical level and the molecular changes are becoming increasingly well-defined (Figure 1.1) (Vogelstein *et al.*, 1988; Fearon and Vogelstein, 1990; Kinzler and Vogelstein, 1996; Boland, 2000). Initiation, promotion and progression of colorectal tumourigenesis are believed to result from a series of genetic alterations leading to progressive dis-regulation of the normal mechanisms controlling growth. Vogelstein *et al*, demonstrated a median of four or five allelic losses in a study of 92 colorectal carcinomas (Vogelstein *et al.*, 1988) and the mutation of at least five genes has been proposed as necessary for the formation of a malignant tumour (Fearon and Vogelstein, 1990). In addition, it appears that the accumulation of changes rather than their order with respect to one another is of primary importance (Fearon and Vogelstein, 1990). However, some alterations are consistently associated with specific stages of the adenoma to carcinoma sequence, such as early loss or mutation of *APC* (Kinzler and Vogelstein, 1996; Boland, 2000). Alterations observed frequently in subsets of CRC include mutations of the *C-K-RAS* oncogene (Vogelstein *et al.*, 1988 Delattre *et al.*, 1989) and mutations or loss via chromosomal deletions, of tumour suppressor genes *p53*, *DCC* and *APC*, on chromosomes 5q, 17p and 18q (Figure 1.1) (Delattre *et al.*, 1989; Fearon and Vogelstein, 1990). These changes are accompanied by aneuploidy in most cases (Fearon and Vogelstein, 1990). However, the timing, the mechanisms by which such mutations arise and the exact role of such changes remains to be fully elucidated.

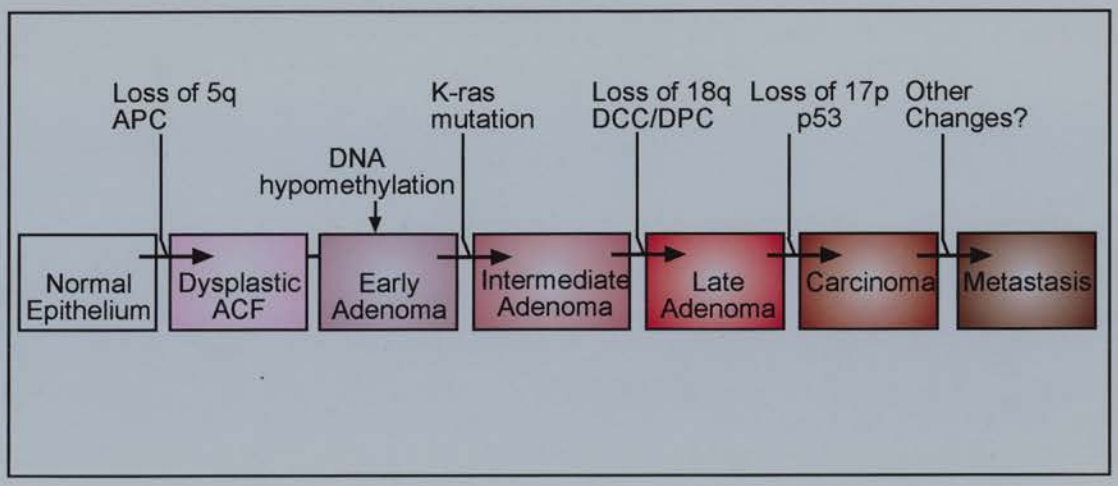


Figure 1.1 Genetic changes associated with the adenoma to carcinoma sequence of colorectal cancer. Tumourigenesis preceeds through a series of genetic changes involving oncogenes (*K-RAS*) and tumour supressor genes such as those on chromosome 5q, 17p and 18q. The clinical stages of this sequence are well defined. Abberent crypt foci (ACF) are thought to be the precursors of adenomas and the three stages of adenomas represent tumours of increasing size, dysplasia and villous content. The overall accumulation of changes seems of greater importance than the order in which they occur. Both selection and instability underlie this process to varying degrees. Adapted from Fearon and Vogelstein, 1990 and Kinzler and Vogelstein, 1996.

1.2.4 Selection and mutation in tumourigenesis

The mechanisms that lead to the high frequency of mutations observed in CRC remains controversial. It is widely accepted that the process of adenoma to carcinoma progression is a continuum, resulting from successive waves of clonal expansion (Nowell, 1976). However, the relative importance of selection for growth advantage and of genomic instability in this process is unclear. Loeb, originally proposed that the expression of a mutator phenotype is an early and necessary step in tumour progression (Loeb *et al.*, 1974). This concept was formulated to explain the disparity between the rarity of mutations in normal cells and the large numbers of mutations, including chromosomal instability, microsatellite instability and anueploidy that are present in a variety of human malignancies such as CRC (Loeb *et al.*, 1974; Peinado *et al.*, 1992; Lengauer *et al.*, 1998; Loeb, 2001). In the "mutator model" of tumourigenesis, initiating mutations are proposed to occur in genes that insure the stability of the genome such as those involved in DNA repair, or those encoding DNA polymerases or helicases (Loeb, 2001). From the population of mutant cells that result consequent of an elevated mutation rate, there is then selection for cells that escape regulatory mechanisms for the control of cell replication, invasion and metastasis. It follows from this model, that genes, which are inherently susceptible to mutation, are more likely to be mutated frequently in cancers. Observations that MMR genes are mutated somatically in a subset of sporadic CRCs and other malignancies, in addition to tumours from HNPCC patients, lends support to the mutator hypothesis (Borresen *et al.*, 1995; Ma *et al.*, 2000). An elevated level of mutation is clearly associated with a variety of familial and non-familial cancers (Merlo *et al.*, 1994; Orth *et al.*, 1994; Kinzler and Vogelstein, 1996).

In disagreement with the mutator hypothesis, Tomlinson and Bodmer have argued that a raised mutation rate in itself does not cause cancer, and that selection is the driving force behind tumour growth (Tomlinson and Bodmer, 1999). They have used mathematical models of tumourigenesis based on CRC to analyse the role of mutation rate in the growth of sporadic tumours. These models demonstrate that selection alone is sufficient and more likely to explain the evolution of tumours, than an increase in intrinsic mutation rate (Tomlinson *et al.*, 1996). They have also found

evidence that mutations in cancer genes such as *APC* occur prior to MMR defects suggesting that an increased mutation rate is not necessary for tumour initiation (Homfray *et al.*, 1998). It has also been suggested that a raised mutation rate may be a disadvantage, particularly in early tumours, since cells with a high mutational load tend to undergo apoptosis (Cahill *et al.*, 1999; Tomlinson and Bodmer, 1999). Under the selection model, the most frequently observed mutations in CRC are those that confer the strongest selectable advantage to the tumour cell. Tomlinson and Bodmer have also highlighted the fact that excision repair mutations rarely occur in sporadic cancers, despite leading to an increased cancer risk via genomic instability in the presence of homozygous germline mutations. This might suggest that there is not a direct association between increased mutation rate and tumourigenesis.

It is likely that both selection and increased mutation contribute to the development of CRC, but the relative balance between these factors in the initiation and progression of colorectal tumourigenesis remains unclear.

Understanding of the genetic consequences of MMR defects may go some way in addressing the relative contribution of genomic instability to the development of CRC.

1.3 Cancers Associated with MMR Defects

1.3.1 Hereditary non-polyposis colorectal cancer (HNPCC)

HNPCC, also known as Lynch Syndrome, was originally called cancer family syndrome after it was noted as a hereditary condition with an autosomal dominant mode of inheritance (Lynch *et al.*, 1966). The first description of such a family (Family G), was made in 1895 by the pathologist Aldred Warthin (Warthin, 1913). Family G was re-studied by Lynch and Krush in 1971 and found to have features of HNPCC (Lynch and Krush, 1971). Within the last ten years the molecular basis of this disorder has been elucidated and has been linked to germline defects in the MMR genes (Lynch and Smyrk, 1996). To date seven human genes that participate in the MMR process have been identified (Fishel *et al.*, 1993; Leach *et al.*, 1993;

Nicolaides *et al.*, 1994; Bronner *et al.*, 1994; Papadopoulos *et al.*, 1994; Konishi *et al.*, 1996). These are *MLH1, MLH3 PMS2, PMS1, MSH2* and *MSH6, MSH3* and germline mutations in at least five of these genes have been found in HNPCC kindreds (Leach *et al.*, 1993; Bronner *et al.*, 1994; Nicolaides *et al.*, 1994; Miyaki *et al.*, 1997a; Kolodner and Marsischky, 1999; Lipkin *et al.*, 2000). Affected family members inherit one mutant allele of the relevant MMR gene and one wild type allele. During the early stages of tumour development, the wild type allele is inactivated through somatic mutation resulting in complete loss of MMR activity. This leads to a mutator phenotype, which can be visualised in tumours as microsatellite instability, in which short repeated sequences are characteristically altered. Inactivation of MMR leads to the rapid accumulation of somatic mutations once tumourigenesis is initiated (Lynch and Smyrk, 1996). Tumour progression is thus accelerated and HNPCC patients consequentially develop CRCs at an average age of 45, more than two decades earlier than the general population (Lynch *et al.*, 1985; Bellacosa *et al.*, 1996; Dunlop, 1997). These genetic mechanisms of HNPCC are detailed below.

An autosomal dominant mode of inheritance and early onset of CRC are two of the clinically defining features of HNPCC. CRC develops in gene mutation carriers without the myriads of adenomas seen in FAP (Lynch *et al.*, 1985). Although, adenomatous polyps are detected in HNPCC patients, these are in numbers comparable to the general population (usually <10) and it is the increase in genomic instability within these adenomas that results in their rapid progression to malignancy (Aaltonen *et al.*, 1994; Lynch and Smyrk, 1996). In HNPCC families there is preferential involvement of the proximal colon with approximately 70% of HNPCC colon cancers versus 30% of sporadic tumours being located proximal to the splenic flexure (Aaltonen *et al.*, 1994; Lynch and Smyrk, 1996). Multiple CRCs may develop and these may be synchronous or metachronous. Pathological features of CRC include poor differentiation, increased signet cells, crohns like reaction, peritumourol lymphocytic infiltration and tumour infiltrating lymphocytes mixed with tumour cells (Jass, 2000). In addition to an increased risk of cancer of the colon and rectum, gene carriers are also at high risk of malignancy at extracolonic sites (Watson and Lynch, 1993; Aaltonen *et al.*, 1994; Bellacosa *et al.*, 1996). These

specifically include the endometrium, ovary, stomach, small bowel, hepatobillary tract, pancreas, ureta and renal pelvis (Watson and Lynch., 1993; Lynch and Smyrk, 1996). An excess of breast cancer has also been noted in some HNPCC families (Risinger *et al.*, 1996). The penetrance for developing HNPCC in gene carriers can approach 100%. Dunlop *et al* reported a risk for gene carriers to age 70 for all cancers of 91% in males and 69% in females (Dunlop *et al.*, 1997). It is of interest that the risk of developing CRC was observed to be significantly greater for males than for females (74% vs 30%) but that the risk of uterine cancer exceeded that for CRC in females (42%) (Dunlop *et al.*, 1997). This suggests that females are protected from CRC in some way, perhaps due to environmental factors or sex linked modifier genes (Dunlop *et al.*, 1997).

Overall the prognosis for individuals from HNPCC families is better than for those with sporadic CRCs. In a Finnish based population study based on patients from the nation-wide Finnish cancer registry, Sankila *et al* found that the five year survival rate for HNPCC patients was 65% compared to 44% for those with sporadic CRC (Sankila *et al.*, 1996).

Since large families are uncommon minimum criteria were drawn up to define HNPCC for research purposes and to establish the frequency of HNPCC in the population, (Vasen *et al.*, 1991). The original criteria known as the Amsterdam Criteria stated that; 1) At least three relatives should have histologically verified CRC; one of them should be a first degree relative to the other two. FAP should be excluded. 2) At least two successive generations should be affected. 3) In one of the relatives CRC should be diagnosed under the age of 50. (Vasen *et al.*, 1991). However, when using these criteria on small families or in cases where little pedigree information is available, there is less likelihood that the definition of HNPCC will be fulfilled and these families may be inappropriately excluded. Furthermore, the Amsterdam criteria require CRC in all affected family members and thus ignore cases where key members are affected by other HNPCC cancers such as those of the endometrium (Vasen *et al.*, 1991; Watson and Lynch, 1993). Evidence from the analysis of the genes responsible for the majority of HNPCC, has therefore lead to the expansion of these criteria to include extra-colonic HNPCC cancers (Vasen *et al.*, 1999). It has been proposed that the presence or absence of a germline mutation in a

MMR gene should also be incorporated into the definition of HNPCC. But this may also be problematic in that not all families fulfilling the Amsterdam criteria are positive for MMR gene mutations (Aaltonen *et al.*, 1994; Liu *et al.*, 1996; Moslein *et al.*, 1996). In an analysis of 74 Amsterdam positive families in which all five MMR genes were analysed, germline mutations were identified in 70% of the kindreds (Liu *et al.*, 1996). Those cases in which MMR mutations have not been identified may be due to failure to detect all mutations as a result of the techniques used. In addition, it is possible that MMR may be inactivated by some other unidentified mechanism or there may be germline defects in as yet undefined genes. The identification of the causative factors in such cases, and further development of rigorous methods of mutation detection will facilitate further systematic assessment of the prevalence of HNPCC on a whole population basis.

1.3.2 Turcots syndrome

In addition to HNPCC, germline defects in MMR contribute to several other rare disorders that are associated with CRC. Turcots syndrome (TS) is characterised by the concurrence of primary brain tumours and multiple adenomas or CRCs that occur at an early age. The malignant transformation of these lesions occurs even earlier than in FAP or HNPCC (Crail, 1949; Turcot *et al.*, 1959) To date over 130 cases of TS have been described but as yet, it has not been established whether the mode of inheritance is dominant or recessive (Matsui *et al.*, 1998; De Rosa *et al.*, 2000).

Clues as to the molecular basis of TS were revealed by analysis of dominantly inherited cases by Hamilton *et al* (Hamilton *et al.*, 1995). In this study, the association between brain tumours and multiple colorectal adenomas was demonstrated to result from two distinct germ line defects; mutation of the *APC* gene or mutation of MMR genes, specifically *PMS2* or *MLH1* (Hamilton *et al.*, 1995). The clinical phenotypes of TS associated with either of these defects are broadly distinguishable. TS individuals with germline *APC* mutations have large numbers of colorectal adenomas characteristic of FAP and brain tumours that are predominantly medulloblastomas (Hamilton *et al.*, 1995). TS individuals with germline defects in the MMR genes tend to have small numbers of colorectal neoplasms as found in

HNPCC, that develop in childhood or early adolescence. In addition they are prone to developing glioblastomas or café au lait spots (Hamilton *et al.*, 1995). However, on purely clinical grounds there appears to be some overlap between the phenotypes associated with *APC* or MMR gene mutations. Hamilton *et al*, reported that in one of their TS families and in the family originally described by Turcot *et al* (Turcot *et al.*, 1959), large numbers of adenomatous polyps were present despite the identification of MMR germline mutations rather than of the *APC* gene (Hamilton *et al.*, 1995).

The tumours of TS patients with germline MMR gene mutations have been demonstrated to show MSI, characteristic of the MMR defects present in HNPCC tumours (Hamilton *et al.*, 1995; De Rosa *et al.*, 2000). In addition a number of reports have also demonstrated MSI within the normal tissue of these patients, a phenotype not generally observed in patients with classical HNPCC (Parsons *et al.*, 1995a; De Rosa *et al.*, 2000). It has been suggested that this high level of DNA instability in the normal tissues, might trigger the early development of cancer in these TS patients (De Rosa *et al.*, 2000). This disease represents an interesting cancer syndrome in which an elevated mutation rate consequent of MMR defects is associated with a clinical phenotype that shares phenotypic characteristics with HNPCC.

1.3.3 Muir-Torre syndrome

Muir-Torre Syndrome (MTS) is an autosomal dominant cancer susceptibility syndrome also associated with defects in the MMR genes. There have been over 150 reported cases of MTS, which is defined by the coincidence of at least one sebaceous skin tumour and one internal malignancy (Muir *et al.*, 1967; Torre, 1968). Although these may occur between the ages of 23 to 90, the average age of presentation is around the age of 50 (Cohen *et al.*, 1991). The skin tumours include sebaceous adenomas, epitheliomas and carcinomas. These sebaceous gland tumours, which are rare in the general population, are considered to be the hallmark of MTS. The spectrum of internal malignancy is similar to that in HNPCC and patients with MTS have also been identified in HNPCC families (Cohen *et al.*, 1991). CRCs are the most common internal malignancy accounting for over 50% of primary cancers,

while endometrial cancers occur with a frequency of about 15% in females (Cohen *et al.*, 1991).

Germline mutation in the MMR genes *MSH2* and *MLH1*, have been identified in some families with this syndrome (Kolodner *et al.*, 1994; Bapat *et al.*, 1996; Kruse *et al.*, 1996; Kruse *et al.*, 1998). Kruse *et al* detected germline mutations in 9/13 (69%) MTS patients ascertained on the basis of sebaceous skin tumours and occurrence of CRCs. All nine of these patients with MMR defects, also exhibited MSI in at least one skin tumour (Kruse *et al.*, 1998). Significantly more mutations were identified in the *MSH2* gene than in *MLH1* and this has also been noted in several other studies (Kolodner *et al.*, 1994; Kruse *et al.*, 1996; Kruse *et al.*, 1998).

It has been suggested that MTS consist of two subgroups. The first being those patients without CRC and the second being those MTS patients with CRC. This second group of patients, who are affected by CRC, may be allelic variants of HNPCC and be associated with defects in MMR. It remains to be determined whether MTS constitutes a distinct genetic entity or whether the observed skin lesions are merely a pleiotropic effect of a MMR deficiency. MTS represents another interesting example in which a raised mutation rate consequent of repair defects is associated with both CRC and diverse extracolonic cancers.

1.3.4 Sporadic cancers

MSI arises in around 13% of sporadic CRCs (Kinzler and Vogelstein, 1996) and these are also associated with defects in MMR genes (Liu *et al.*, 1995a; Liu *et al.*, 1995b). In a more recent prospective study of 257 unselected CRC patients, 20% were found to have MMR defects with 2% being accounted for by familial cases (Cunningham *et al.*, 2001). This indicates that the frequency of defective DNA MMR in CRC is relatively common.

Since neither MMR gene allele is mutated in the germline of sporadic CRCs, inactivation of both alleles occurs somatically and mutation of *MLH1, MLH3 MSH2, MSH3, MSH6* and *PMS2* has been demonstrated in sporadic cases of CRC (Liu *et al.*, 1995b; Malkhosyan *et al.*, 1996; Lipkin *et al.*, 2000; Ma *et al.*, 2000).

Sporadic CRCs with MMR defects share certain features with CRCs from HNPCC patients (Liu *et al.*, 1995b). Both tend to be right sided and relatively diploid, unlike MMR proficient CRCs that usually have an anueploid phenotype. In addition, both have a relatively good prognosis. However, while HNPCC patients often manifest left sided CRCs and extra-colonic cancers sporadic cases of CRC almost always occurs singly and in the proximal colon (Aaltonen *et al.*, 1994; Liu *et al.*, 1995b). In addition, the age of onset of sporadic CRCs with MMR defects is later than in HNPCC patients by more than 20 years (Liu *et al.*, 1995b). Interestingly, there is a subset of CRC patients who develop the disease at a very young age (under 35 years) but do not have a family history of HNPCC and the majority of these have MMR defects (Liu *et al.*, 1995a).

MSI has been identified in a variable fraction of several other types of neoplasm's including those of the lung, breast, pancreas, endometrium and ovary (Risinger *et al.*, 1993; Han *et al.*, 1993; Orth *et al.*, 1994; Merlo *et al.*, 1994; Gurin *et al.*, 1999; Caldes *et al.*, 2000; Ottini *et al.*, 2000). MMR defects have also been identified in some of these cases (Orth *et al.*, 1994; Chung *et al.*, 1997). These reports provide evidence that an increased mutation rate consequent of MMR defects is associated with a wide spectrum of primary malignancies.

1.4 Mismatch Repair

Elucidating the mechanism of MMR has proved essential in understanding how its dys-regulation contributes to the accumulation of mutations during tumourigenesis. MMR is critical for maintaining the overall integrity of the genome and enhances the fidelity of chromosome replication by 100 to 1000 times (Modrich, 1991). The basic features of the MMR system have been highly conserved during evolution and homologues of the bacterial MutS and MutL proteins have been identified in yeast, mammalian and other eukaryotic cells (Harfe and Jinks-Robertson, 2000). These proteins play key roles in mismatch recognition and initiation of repair. The primary function of the MMR system is to eliminate base:base mismatches and insertion:deletion loops. Such lesions arise in DNA by

physical damage to existing nucleotides, polymerase mis-incorporation errors during DNA replication, and as a result of forming heteroduplex intermediates during the process of genetic recombination (Freidberg *et al.*, 1995). Failure to repair base:base mis-pairs typically affects non-repetitive DNA and leads to single base substitutions (e.g. G→T). Failed correction of insertion:deletion loops affects repetitive DNA, resulting in the gains or losses of short repeat units within microsatellites manifesting as a characteristic phenotype known as MSI.

The mechanism of MMR comprises several universal features. Initially, error correction involves the efficient recognition of mismatches. This is followed by the selective removal of the newly synthesised strand containing the incorrect information and then re-synthesis of the correct strand (Harfe and Jinks-Robertson, 2000). Strand discrimination is an essential feature of all MMR systems since it ensures the incorrect strand is not used as the template. However, this latter feature appears to be less well conserved between MMR systems in prokaryotes and eukaryotes.

1.4.1 Prokaryotic mismatch repair

Since the key proteins in MMR are highly conserved from bacteria to mammals, much of what is understood about MMR mechanisms in eukaryotes is based on studies of the prokaryotic system. In prokaryotes the best understood MMR pathway is the DNA adenine methylase (DAM) instructed MutHLS pathway in *E.coli* also known as long patch repair (Modrich, 1991). This pathway is responsible for the repair of most nucleotide mispairs and has been studied extensively in vitro and in vivo (Modrich, 1991). The idea that this pathway suppresses misincorporation error is supported by the observation that *mutS, mutL, mutH, mutU* (*uvrD*) and *dam* mutations result in *E. coli* strains with elevated frequencies of spontaneous mutation (Glickman and Radman, 1980). Characterisation of the functions of the *mut* genes in correcting replication errors was facilitated by the purification of the protein products and in vitro reconstruction of the *E.coli* MMR reaction (Lahue *et al.*, 1989; Modrich and Lahue, 1996).

Briefly, MutS is an ATPase that drives mismatch recognition. MutL is also an ATPase, that couples mismatch recognition to MutS by downstream processing steps and MutH is involved in targeting MMR to the newly synthesised strand, being a methylation sensitive endonuclease (Modrich and Lahue, 1996) (Figure 1.2).

MutS is a 95kD ATPase that acts as a homodimer to bind base:base mismatches or small insertion:deletion loops (Su and Modrich, 1986). Almost all mismatches are recognised by MutS and subsequently repaired, although there is some variation in the affinity and efficiency of repair depending on the nature of the mismatch and the sequence context (Su and Modrich, 1986; Fishel and Kolodner, 1995; Modrich and Lahue, 1996). The elucidation of the crystal structure of the *E.coli* and *T. aquaticus* MutS homodimers each bound to a mismatch, has provided considerable insight into the function of this protein (Lamers *et al.*, 2000; Obmolova *et al.*, 2000). Both reports show that the overall structure of the MutS dimer is asymmetric with the two MutS monomers assuming different conformations (Lamers *et al.*, 2000; Obmolova *et al.*, 2000). This finding is significant as it explains why the eukaryotic MutS complexes (described below) exist only as heterodimers. The crystal structure also reveals that the mismatch is specifically recognised by just one of the monomers, which is ADP bound (Lamers *et al.*, 2000). However both monomers contact the DNA forming a clamp-like structure (Lamers *et al.*, 2000; Obmolova *et al.*, 2000).

The strand discrimination signal in the *E.coli* MMR system is provided by the transiently unmethylated state of the newly synthesised DNA. The role of the 25kD endonuclease MutH protein, is to cleave the unmethylated strand of a hemi-methylated GATC dam methylation site thereby introducing a nick in the nascent strand for exonucleolytic removal and re-synthesis (Modrich, 1991; Au *et al.*, 1992). The distance separating the strand signal and the mismatch can be substantial. The recognition d(GATC) site, can direct mismatch correction from a kilobase away but the strength of this signal is greatly reduced when separation distances exceed 2kb (Modrich and Lahue, 1996). However, the *E. coli* system of strand discrimination is not universal among prokaryotes and is not conserved in eukaryotic systems (Modrich, 1991). It has been demonstrated that in the absence of MutH activity a pre-existing nick on one strand of a duplex is sufficient to confer strand specific

repair in vitro and this applies to eukaryotic systems (Lahue *et al.*, 1989; Harfe and Jinks-Robertson, 2000).

The binding of MutS to a mismatch is followed, by ATP dependent binding of the ATPase MutL homodimer (Grilley *et al.*, 1989; Ban and Yang, 1998). This results in an enhancement of bi-directional scanning until a strand discrimination signal is encountered. Although genetic studies have demonstrated that the 70kD MutL protein is essential for MMR, its precise role is less well defined than that of MutS and MutH (Lahue *et al.*, 1989). However a number of studies have indicated that it serves as a molecular matchmaker, coupling mismatch recognition and downstream MMR events. MutL homodimers have been demonstrated to form a complex with MutS (Grilley *et al.*, 1989; Galio *et al.*, 1999) and enhance ATP hydrolysis (Modrich and Lahue, 1996) and this is probably as part of the identification of the strand discrimination signal. MutL then stimulates MutH endonuclease in an ATP dependent manner (Ban and Yang, 1998; Hall and Matson, 1999). Finally MutL is required to load MutU (Uvrd) at the site of the MutH induced nick, thus facilitating DNA unwinding which proceeds towards the mismatch and directs subsequent removal of the nascent strand (Modrich and Lahue, 1996).

Proteins involved in general DNA metabolic processes complete the latter steps of MMR. The excision step of MMR appears to require any one of four single stranded DNA specific exonucleases, RecJ, ExoVII, ExoI, or ExoX since MMR is inactivated only when all four exonucleases are inactivated both in vitro and in vivo (Modrich, 1991). Polymerase II holoenzyme and DNA ligase complete re-synthesis and ligation (Lahue *et al.*, 1989)

In addition to the long patch repair system, E.coli possesses a second short patch system that shares some components with the MutHLS system (Harfe and Jinks-Robertson, 2000). Furthermore a third system, the very short patch (VSP) system, specifically repairs G/T mismatches to G/C in the context of DNA cytosine methytransferase (DCM), and its primary role is thought to be in correction of mismatches resulting from deamination of 5-methylcytosine to thymine (Modrich, 1991).

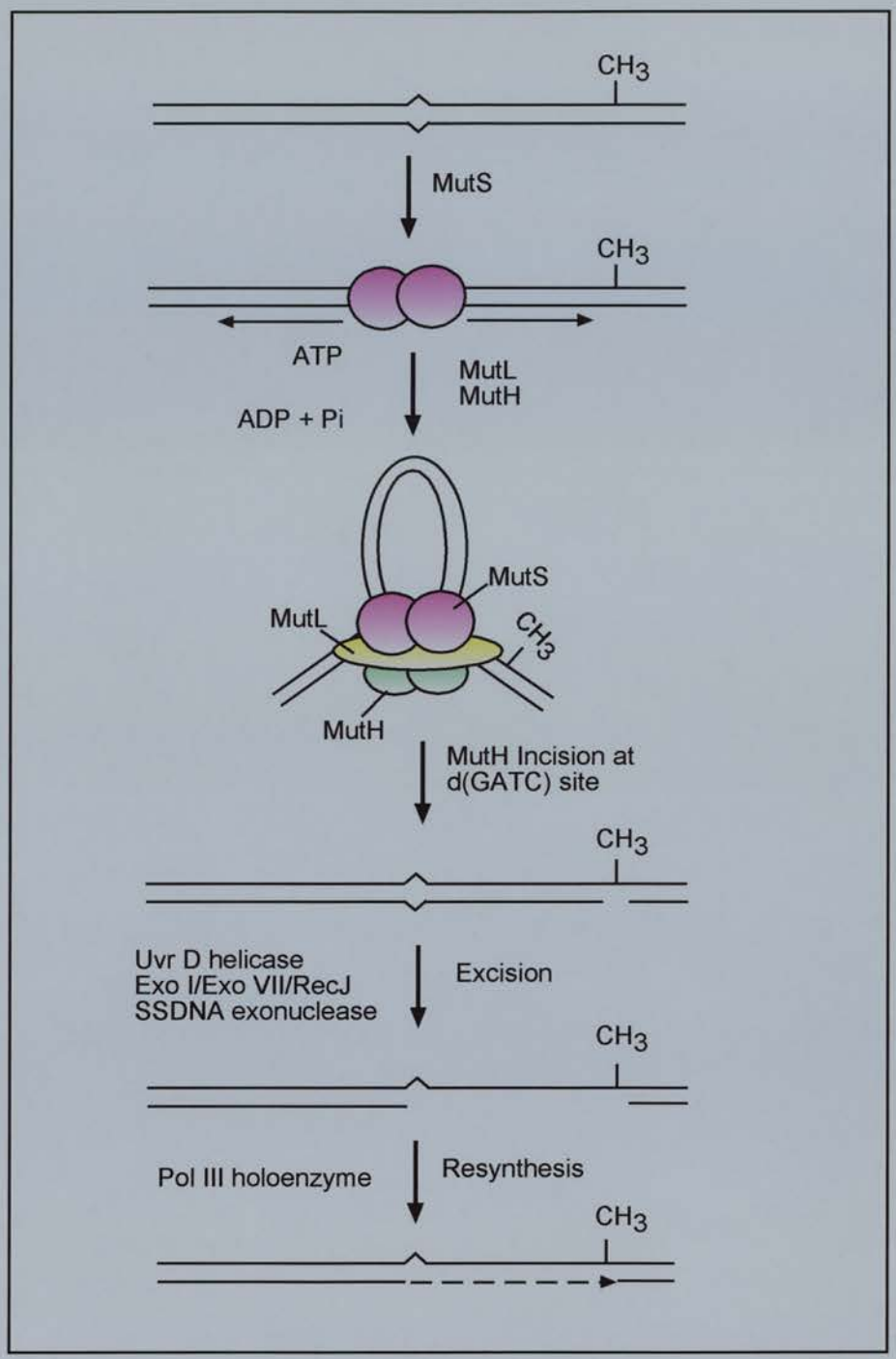


Figure 1.2 Model for mismatch repair of DNA replication errors in *E.coli*. Adapted from Modrich, 1991, Fishel and Kolodner, 1995 and Harfe and Jinks-Robertson, 2000.

1.4.2 Eukaryotic mismatch repair

The function of the eukaryotic MMR system is best understood in the budding yeast *S. cerevisiae* and human MMR appears to share genetic and functional similarities with this system. Like bacteria, mutation of the *S.cerevisiae* MMR genes was observed to result in yeast strains with mutator phenotypes, providing considerable insight into the specific roles of the MMR proteins (Marsischky *et al.*, 1996).

All eukaryotic organisms characterised to date posses multiple MutS homologues (MSH proteins) and multiple MutL homologues (MLH proteins) and the yeast and human homologues are shown in Table 1.1. However, convincing MutH homologues have yet to be identified and there is no evidence to suggest that strand discrimination is methyl-directed (Harfe and Jinks-Robertson, 2000).

Table 1.1 Eukaryotic MutS and MutL homologues. Adapted from Harfe and Jinks-Robertson, (2000). The human nomenclature is used and primary functions are shown based on data from both yeast and mammalian cells. NI indicates not identified

E.coli	*S.cerevisiae*	Humans	Eukaryotic Primary Function
MutS	MSH1	NI	Mutation avoidance in mitochondria
	MSH2	MSH2	Hetrodimerises with MSH3 and MSH6 to:
			Repair base:base mispairs and insertion:deletion loops
			Responds to DNA damage (humans)
	MSH3	MSH3	Heterodimerises with MSH2 to form MutSβ
	MSH4	MSH4	Hetrodimerises with MSH5 and promotes crossing over in meiosis
	MSH5	MSH5	Heterodimerises with MSH4
	MSH6	MSH6	Heterodimerises with MSH2 to form MutSα
MutL	PMS1	PMS2	Hetrodimerises with MLH1 to form MutLα and:
			Repairs base:base mispairs and insertion:deletion loops
			Responds to DNA damage? (humans)
	MLH1	MLH1	Heterodimerises with PMS2 (humans), PMS1 (humans) and MLH3
	MLH2	PMS1	Heterodimerises with MLH1 to form MutLβ and:
			Repairs insertion:deletion loops
	MLH3	MLH3	Heterodimerises with MLH1 to:
			Repair insertion:deletion loops

Yeast possess six homologues (MSH1-6) of the bacterial MutS protein and with the exception of the mitochondrial protein MSH1, human homologues of all the yeast MSH genes have been identified (Table1.1). These are named according to the yeast designations (Harfe and Jinks-Robertson, 2000). In *S.cerevisae* MSH4 and MSH5 are not involved in MMR but are required for crossing over during meiotic recombination (Kolodner and Marsischky, 1999). Consistent with their role in yeast, human MSH4 and MSH5 are most highly expressed in meiotic tissue (Kolodner and Marsischky, 1999).

Hetrodimers formed between yeast MSH2 and either MSH3 or MSH6 recognise mis-paired bases in DNA (Acharya *et al.*, 1996). In humans these complexes have been termed MutSβ and MutSα respectively and their recognition properties are

largely similar to the corresponding *S.cerevisae* complexes (Acharya *et al.*, 1996; Genschel *et al.*, 1998). Marsischky *et al* measured the rate of accumulating mutations and mutation spectrum in yeast strains containing different combinations of *MSH2*, *MSH3* and *MSH6* mutations and suggested that two pathways of MSH2 dependent mismatch repair exist (Marsischky *et al.*, 1996). This has been confirmed by the identification of the recognition specificities of the two MSH complexes in both *S. cerevisiae* and humans (Acharya *et al.*, 1996; Alani, 1996; Johnson *et al.*, 1996; Palombo *et al.*, 1996; Genschel *et al.*, 1998; Umar *et al.*, 1998). These studies have also revealed functional redundancy between MSH3 and MSH6 (Acharya *et al.*, 1996; Johnson *et al.*, 1996; Marsischky *et al.*, 1996; Palombo *et al.*, 1996; Genschel *et al.*, 1998; Umar *et al.*, 1998). While MSH2-MSH6 binds preferentially to base substitutions and small insertion:deletion mismatches, MSH2-MSH3 is specific for insertion:deletion mismatches (Figure 1.3) (Alani, 1996; Habraken *et al.*, 1996; Johnson *et al.*, 1996; Palombo *et al.*, 1996; Umar *et al.*, 1998). The redundancy between MSH3 and MSH6 in the repair of insertion:deletion loops is suggested to explain the prevalence of MSH2 mutations in HNPCC CRCs over those in the other two MSH proteins and also as to why germline mutations in MSH6 have been detected in patients whose tumours exhibit a mild form MSI.

The in vitro interaction between the human MutSα complex with mismatch containing duplexes has been characterised in detail and there are two predominating models to explain the structural changes elicited by ADP-ATP exchange (Gradia *et al.*, 1997; Blackwell *et al.*, 1998). The first model suggests that the MSH complex functions as a nucleotide regulated molecular switch and that ADP and ATP have opposing effects in MutSα mis-pair binding (Gradia *et al.*, 1997). MutSα complexed with ADP is proposed to bind to the mispair with near irreversible affinity. This tight binding acts as a flag for the assembly of the excision repair machinery and subsequent exchange of ADP-ATP results in the release of MutSα from the DNA allowing excision and re-synthesis to take place. On release, the intrinsic ATPase of MutSα hydrolyses ATP resulting in a form that is again competent for mispair binding (Gradia *et al.*, 1997). This is substantially different from the translocation model proposed by Modrich *et al*, that invokes the notion of ATP binding and hydrolysis in movement away from the mismatch containing site (Blackwell *et al.*,

1998). This study demonstrates that ADP is not required for mismatch recognition and that dissociation of MutSα occurs at DNA ends in a reaction dependent on DNA hydrolysis (Blackwell *et al.*, 1998).

Four MutL homologues have been identified in both yeast and humans. In yeast these are PMS1, MLH1, MLH2 and MLH3 however in humans the same yeast names are not applied and this can be a source of confusion (Table 1.1). In both *S. cerevisae* and humans MLH1 is the major MutL homologue and has been demonstrated to form heterodimers with the other three MutL proteins (Prolla *et al.*, 1994; Li and Modrich, 1995; Raschle *et al.*, 1999; Lipkin *et al.*, 2000). PMS1 in yeast is homologous to PMS2 in humans. In yeast, the MLH1-PMS1 complex is proposed to act as a molecular matchmaker, co-ordinating activities between the MSH complexes and downstream MMR events in a role similar to that of MutL in bacteria (Prolla *et al.*, 1994). Interactions between yeast MLH1-PMS1 (MLH1-PMS2 in humans) and MSH2 complexes have been demonstrated in both humans and yeast supporting the role of this MLH complex as a molecular matchmaker (Figure 1.3) (Kolodner and Marsischky, 1999). In humans the MLH1-PMS2 complex is known as MutLα and has been shown to restore repair proficiency on heteroduplex DNAs containing base:base mispairs or insertion:deletion loops, in cell extracts that are deficient for MLH1 (Li and Modrich, 1995). Human PMS1 exhibits homology to both yeast MLH2 and MLH3 and mammalian MLH1-PMS1 hetererodimers are referred to as MutLβ (Raschle *et al.*, 1999). Although the role of MutLβ in MMR remains to be demonstrated, the fact that few HNPCC families have been identified with either PMS1 or PMS2 mutations has lead to the suggestion that there may be partial redundancy in function between these two proteins (Kolodner and Marsischky, 1999; Raschle *et al.*, 1999; Harfe and Jinks-Robertson, 2000). The human MLH3 protein has been more recently identified and demonstrated to interact with MLH1 (Lipkin *et al.*, 2000). Human MLH3 shows most homology to yeast MLH3 which appears to be involved in the repair of a subset of insertion:deletion mispairs recognised by the MSH2-MSH3 complex (Figure 1.3) (Flores-Rozas and Kolodner, 1998). The role of the human MLH1-MLH3 complex remains to be elucidated, although over expression of an MLH3 N-terminal deletion protein is associated with an MSI phenotype (Lipkin *et al.*, 2000).

The assembly of the MLH and MSH complexes subsequently recruits a number of proteins that are not specific to MMR activity but are involved in completing the later stages of MMR. These include DNA polymerase δ, replication protein A, proliferating cell nuclear antigen (PCNA), replication factor C (RFC), exonuclease 1, Fen1 and the DNA polymerase δ and ε associated exonucleases (Prolla, 1998a; Kolodner and Marsischky, 1999).

The human MED1 protein, also known as MBD4 has been speculated to represent a bacterial MutH functional homologue on the basis that it interacts with MLH1 and displays endonuclease activity (Bellacosa *et al.*, 1999). Frameshift mutations have also been identified in MED1 in colorectal tumours exhibiting MSI (Bader *et al.*, 1999; Riccio *et al.*, 1999). However, although MBD4 binds to methyl-CpG-containing DNA, it has greater affinity for fully methylated rather than hemimethylated DNA and there is no evidence that it is involved in strand discrimination (Bellacosa *et al.*, 1999; Hendrich *et al.*, 1999). Furthermore, in the second study by Hendrich *et al*, it was revealed that MBD4 is a mismatch specific T/U glycosylase and no functional evidence for the endonuclease activity reported by Bellacosa *et al* was identified (Hendrich *et al.*, 1999). Rather than being a homologue of bacterial MutH, MBD4 appears to function in the minimisation the mutability at methyl CpG sites (Hendrich *et al.*, 1999). In any case it is widely accepted that strand discrimination in eukaryotes is nick directed as can be the case in *E.coli* (Lahue *et al.*, 1989).

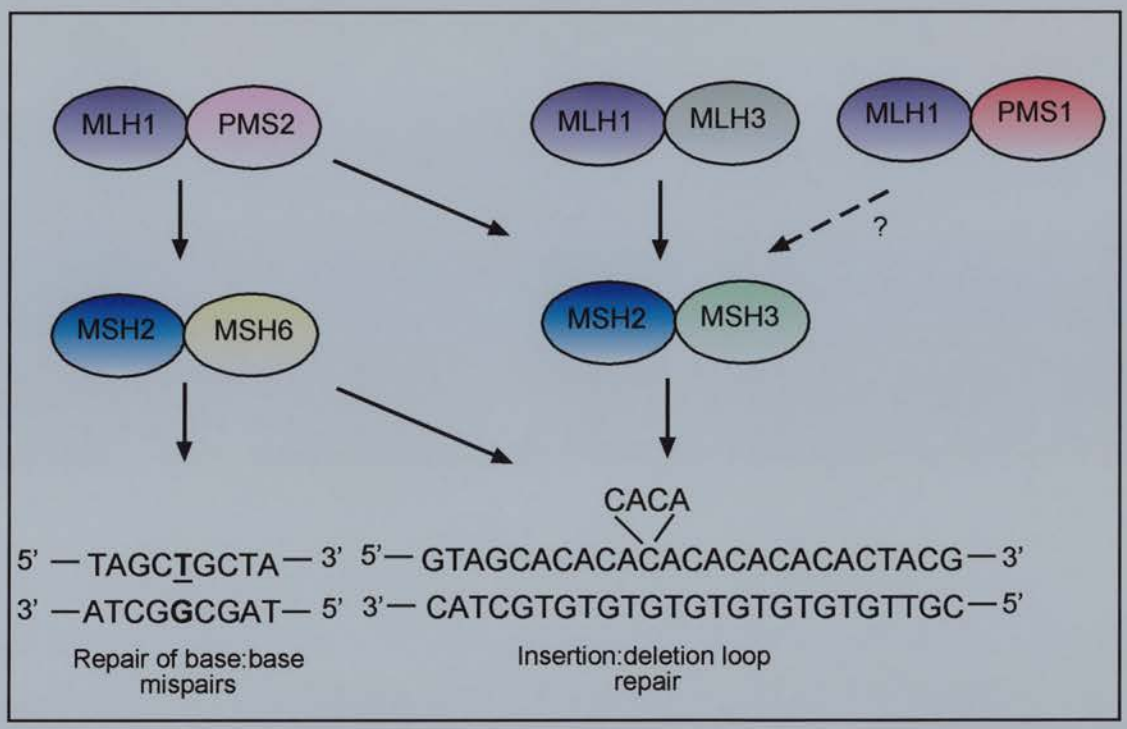


Figure 1.3 Interaction of human MMR proteins in the repair of base:base mispairs and insertion: deletion loops. Adapted from Kolodner and Marsischky, 1999 and Harfe and Jinks-Robertson, 2000.

1.4.3 Role of MMR in the response to DNA damage

The MMR proteins have an additional function in mediating the activation of cell cycle checkpoints and apoptosis in response to DNA damage (Fink *et al.*, 1998; Harfe and Jinks-Robertson, 2000). The elucidation of this role, came from the discovery that MMR defective bacteria are tolerant to alkylating agents (Karran and Marinus, 1982) and this phenomenon was subsequently found to be a characteristic of mammalian cells (Fink *et al.*, 1998). Interestingly yeast MMR proteins do not appear to be involved in response to DNA damage.

Alkylating agents such as N-methyl-N'-nitro-N-nitrosanguanidine (MNNG) and N-methyl-N-nitosourea (MNU) generate a variety of adducts in DNA of which O^6-methyl-guanine (O^6-MeG) is the most cytotoxic. In mammalian cells treatment with MNNG triggers an MMR-dependent G2-M arrest (Hawn *et al.*, 1995), followed by a MMR dependent apoptotic response (Wu *et al.*, 1999a; Zhang *et al.*, 1999). In MMR proficient cells erroneous incorporation of a thymine opposite O^6-MeG is thought to initiate futile cycling of MMR, since the thymine containing strand rather than the O^6-MeG containing strand is excised and the O^6-MeG lesion persists. Continued rounds of MMR at the O^6-MeG lesion, increase the risk of a double strand break at the time of the next S phase, thus inducing apoptosis (Fink *et al.*, 1998). In addition to O^6-MeG lesions, MMR proteins are responsible for triggering apoptosis in response to a wide variety of DNA damaging agents (Fink *et al.*, 1998).

It has been shown that MutSα and MutLβ are specifically required for signalling the initiation of apoptosis in response to alkylating agents (Duckett *et al.*, 1999; Hickman and Samson, 1999). However, Zhang *et al* demonstrated that over expression of MSH2 or MLH1 induced apoptosis in either repair proficient or deficient cells, whereas over-expression of MSH3, MSH6 and PMS2 did not (Zhang *et al.*, 1999).

The precise mechanism by which MMR triggers apoptosis has not been elucidated, although both p53 dependent and p53 independent mechanisms have been demonstrated (Duckett *et al.*, 1999; Hickman and Samson, 1999). It has been suggested that MMR components may be involved in sensing DNA damage and

transducing the damage signal to downstream signalling components. Indeed, a large multisubunit complex termed the BRCA1-associated genome surveillance complex (BASC) has been identified (Wang *et al.*, 2000). Using immunoprecipitation and mass spectrometry, Wang *et al* partially purified a complex containing BRCA1 and a variety of other components all of which are consistent with a function for this complex, in sensing DNA damage (Wang *et al.*, 2000). These components include the MMR proteins MSH2, MSH6 and MLH1.

The identification of a role for MMR in the response to DNA damage has clinical implications in terms of the chemotherapeutic treatment of MMR deficient cancers. Chemotherapeutic agents induce DNA damage specifically to trigger cell death thus destroying cancerous cells. However, MMR deficient cells that are no longer responsive to damage, acquire resistance to this type of treatment (Fink *et al.*, 1998).

1.5 MMR Gene Inactivation

1.5.1 Identification of the MMR genes

Although HNPCC has been recognised as a hereditary condition for over 25 years, the molecular basis of this disease has only been determined in the last eight. Of the six major human MMR genes that have been identified (*MSH2, MLH1, MLH3, PMS1, PMS2, MSH6, MSH3*), germline and somatic mutations in all but *MSH3* and *MLH3* have been associated with HNPCC and a subset of sporadic CRCs (Fishel *et al.*, 1993; Leach *et al.*, 1993; Bronner *et al.*, 1994; Nicolaides *et al.*, 1994; Papadopoulos *et al.*, 1995; Miyaki *et al.*, 1997a; Loukola *et al.*, 2000).

The genetic basis for HNPCC was initially reported by Peltomaki *et al* in 1993 (Peltomaki *et al.*, 1993). In this study, a panel of 345 microsatellite markers, covering the whole genome was used in linkage analysis of two families who met the HNPCC criteria (Vasen *et al.*, 1991). Highly significant linkage to marker D2S123 led to the mapping of the first HNPCC locus at 2p15-16 (Peltomaki *et al.*, 1993). This was followed by the identification of a second HNPCC locus by Lindblom *et al* who combined restriction fragment length polymorphisms with microsatellite

markers to link cancer occurrences with markers on chromosome 3p (Lindblom *et al.*, 1993).

Fishel *et al* cloned the first human MMR gene located at the HNPCC locus on chromosome 2 (Fishel *et al.*, 1993). *MSH2* was localised to 2p21-22 and mutations were identified in sporadic tumours and within the germline of individuals from two HNPCC kindreds (Fishel *et al.*, 1993). A similar observation was made in a second study (Leach *et al.*, 1993). The knowledge that multiple genes function in MMR, indicated that HNPCC might be associated with defects in any one of them. In 1994 the *mutL* homologue *MLH1* was cloned and located to the second HNPCC locus at 3p21 (Bronner *et al.*, 1994). PMS1 and PMS2 were also identified, by homology to the bacterial and yeast proteins and were localised by fluorescent in situ hybridisation to 2q31-33 and 7p22 respectively (Nicolaides *et al.*, 1994). Mutations in both genes were identified in the germline of HNPCC patients (Nicolaides *et al.*, 1994). MSH6, originally named GTBP for G/T binding protein, was localised to 2p16 and demonstrated to function in MMR (Palombo *et al.*, 1995; Papadopoulos *et al.*, 1995). However, the unusual phenotype associated with *MSH6* mutations meant that it was several more years before germline mutations in this gene were identified in HNPCC families (Miyaki *et al.*, 1997a).

The MMR genes are generally accepted to function as tumour suppressors in that heterozygous cells have normal repair activity and loss or mutations of the wild type allele is necessary for inactivation of MMR and subsequent tumour development. In the inherited MMR defective disorders, one allele is already mutated in the germline and somatic inactivation of the remaining wild type allele must take place. In sporadic cases both alleles are intact in the germline and thus two somatic inactivation events must occur.

1.5.2 Germline mutations of the MMR genes

A database of HNPCC associated mutations is maintained at http://www.nfdht.nl which to date, contains information on over 300 germline mutations occurring in over 500 HNPCC families from all parts of the world (Peltomaki, 2001). Around

50% of germline mutations occur in *MLH1* and 40% in *MSH2* (Peltomaki, 2001) (www.nfdht.nl). The predilection for mutations in these two genes is due to the fact that loss of either results in the complete inactivation of base:base mispairs and insertions:deletion loops as described (Figure 1.3). As a consequence, a strong mutator phenotype and high levels of MSI are observed in the tumours of *MSH2* and *MLH1* kindred's (Liu, 1996). Mutations in *MLH1* and *MSH2* are quite evenly distributed although some clustering has been observed in exon 12 of *MSH2* and exon 16 of *MLH1* (Peltomaki and Vasen, 1997). In the 1996 version of the mutation database, Peltomaki noted that the mutations in *MLH1* were mainly frameshift (40%) or missense (31%) alterations, whereas germline changes in *MSH2* were mainly frameshifts (60%) or non-sense (23%) mutations (Peltomaki and Vasen, 1997).

Most of the reported mutations are unique with no real detectable hotspots. However, there have been some notable cases of founder mutations that account for a number of HNPCC families. One example is the exon 16 deletion of *MLH1* known as the Finland 1 mutation. It occurs in around 40 unrelated families in Finland and Sweden. Genealogical studies have demonstrated that in a geographical cluster in south central Finland, most of the affected individuals carrying this mutation belong to different families descended from an ancestor who was one of a small number of "founders" of this Finnish subpopulation around 500 years ago (Lathi *et al.*, 1994).

Due to the functional redundancy between human MutSα and MutSβ in the repair of insertion:deletion mispairs (Figure 1.3), *MSH3* mutations would not be expected to result in a mutator phenotype. Consistent with this, no *MSH3* kindred's have been reported to date (www.nfdht.nl).

The number of HNPCC families in which germline *MSH6* mutations have been identified, has rapidly increased over the last five years (Peltomaki and Vasen, 1997; Peltomaki, 2001). Due to the functional redundancy between MSH6 and MSH3, *MSH6* mutations predispose to an atypical form of HNPCC and this lead to *MSH6* kindreds being overlooked in initial studies (Akiyama *et al.*, 1997b). Inactivation of MSH6, results in high rates of accumulation of base substitution mutations, since the MSH2-MSH3 complex can still efficiently repair insertion:deletion mispairs. Dinucleotide repeat markers were initially used to examine HNPCC families who

showed the characteristic MSI^+ phenotype. Loss of MSH6 would not result in insertions or deletions at such loci and thus exclusion of potential *MSH6* mutant families from study resulted. The tumours of patients with *MSH6* mutations typically have a weaker MSI phenotype compared to the tumours from patients with *MLH1* or *MSH2* defects (Kolodner *et al.*, 1999). It has been suggested that this reduced level of MSI may be the cause of the late onset, low penetrence form of HNPCC that typifies patients with germline mutations of *MSH6* (Akiyama *et al.*, 1997b; Kolodner *et al.*, 1999).

Despite the fact that the numbers of families with mutations in *MSH2*, *MLH1* and *MSH6* is steadily growing only five *PMS2* germline mutations have been detected and a single incidence of an HNPCC family with a germline *PMS1* mutation identified (Nicolaides *et al.*, 1994) (www.nfdht.nl). The involvement of *PMS1* mutations in a cancer family is intriguing in itself since the protein apparently plays no role in MMR (Raschle *et al.*, 1999). Furthermore, of the five *PMS2* mutations detected, only two occurred in classical HNPCC families (Nicolaides *et al.*, 1994). The remaining mutations were discovered in patients with familial CRC but also phenotypic evidence of TS (Hamilton *et al.*, 1995; Miyaki *et al.*, 1997b; De Rosa *et al.*, 2000). This may indicate that germline mutations in *PMS2* predispose individuals to more atypical forms of HNPCC, as is the case for mutations of *MSH6.*

1.5.3 MMR defects in mice

The generation of MMR defective mice has provided significant insight into the role of the MMR proteins in the development of cancer, as well as in the mechanisms of repair itself (Heyer *et al.*, 1999).

Msh2 was the first MMR gene to be inactivated in the mouse. (De Wind *et al.*, 1995; Reitmair *et al.*, 1995). Although $Msh2^{-/-}$ mice are fertile (Reitmair *et al.*, 1995), they have a significantly decreased survival compared to both $Msh2^{+/+}$ and $Msh2^{+/-}$ mice. Around 50% are dead by 6 months and 100% die by one year of age (Reitmair *et al.*, 1996). Interestingly these mice do not develop early-onset CRC, as do HNPCC individuals, but are predisposed to lymphomas early in life. (Reitmair *et al.*, 1995;

Reitmair *et al.*, 1996). However, mice surviving beyond 6 months do develop gastrointestinal tumours and these demonstrate MSI, characteristic of defects in MMR. In addition, MSI was also detected in subclones from embryonic fibroblast cell lines established from *Msh2*$^{-/-}$ mice (Reitmair, *et al.*, 1997). A small percentage of Msh2 deficient mice also develop skin neoplasms, similar to the type occurring in MT syndrome (Reitmair *et al.*, 1996).

Mlh1$^{-/-}$ mice show a similarly reduced survival to *Msh2*$^{-/-}$ mice. Like Msh2 deficient mice, Mlh1 deficient mice have a pre-disposition predominantly for lymphomas and also for GI tumours and again MSI is detected in these tumours (Prolla *et al.*, 1998b; Edelmann *et al.*, 1999). The fact that *Mlh1*$^{-/-}$ mice are infertile supports other lines of evidence that suggest MLH1 has a role in meiosis (Edelmann *et al.*, 1996).

The MMR mutations in both *Msh2*$^{-/-}$ and *Mlh1*$^{-/-}$ mice are clearly compatible with normal development, but loss of MMR function leads to a predisposition for cancer that is manifested by lymphoid tumours. Although a large number of tissue types are involved in the tumours of HNPCC patients, lymphoid tumours are infrequent. However, they have been observed in cases from MT and TS families, both of which are associated with MMR defects (Hamilton *et al.*, 1995; Kruse *et al.*, 1998). The propensity for lymphomas in Msh2 and Mlh1 deficient mice, may reflect intrinsic differences between mice and humans to lymphoma susceptibility. Alternatively, the high turnover of maturing T-cells that exists in the developing immune system of new-born mice, may offer a window in which MMR deficiency can strongly accelerate the accumulation of transforming events (De Wind *et al.*, 1995). This explanation is indicative of a raised mutation rate having a major role in carcinogenesis within expanding cell populations.

The idea that defects in *MSH6* lead to an atypical form of HNPCC is supported by the phenotype of the *Msh6*$^{-/-}$ mouse (Edelmann *et al.*, 1997). While these mice developed B and T- cell lymphomas and GI cancers, these tumours do not display MSI. Extracts of *Msh2*$^{-/-}$ ES cells were demonstrated to be completely defective in the repair of base:base mispairs, but not in the repair of insertion:deletion mispairs consistent with the predicted redundancy between human and yeast MSH6 and

MSH3 (Edelmann *et al.*, 1997). In addition the life span of these animals, while reduced, is longer than that observed for *Msh2*$^{-/-}$ and *Mlh1*$^{-/-}$ mice. By 10-11 months 50% are dead and 100% die by 18 months (Edelmann *et al.*, 1997). Generation of the *Msh6*$^{-/-}$ mouse has provided evidence that the severity of the cancer pre-disposition reflects the severity of the causative MMR defect. That MSH3 may co-operate with MSH6 in tumour suppression was demonstrated by the generation of a double Msh6$^{-/-}$*Msh3*$^{-/-}$ mouse (Edelmann *et al.*, 2000). In contrast to the atypical phenotype of the *Msh6*$^{-/-}$ mouse, the double mutant was indistinguishable in terms of tumour predisposition phenotype to *Mlh1*$^{-/-}$ and *Msh2*$^{-/-}$ mice (Edelmann *et al.*, 2000).

Mice deficient in Pms1 are not prone to cancer supporting the idea that human PMS1 is either not involved in MMR or shares redundancy with another MutL homologue, possibly MLH3 (Prolla *et al.*, 1998b). The lack of cancers in the *Pms1*$^{-/-}$ mouse makes the observation of an HNPCC family with a germline *PMS1* mutation even more intriguing (Nicolaides *et al.*, 1994).

Consistent with the observation of *PMS2* mutations in HNPCC kindreds (Nicolaides *et al.*, 1994), mice deficient in PMS2 are prone to cancer. *Pms2*$^{-/-}$ mice are pre-disposed to lymphomas like *Msh2*$^{-/-}$ and *Mlh1*$^{-/-}$ mice, but they do not develop GI lesions (Prolla *et al.*, 1998b). Furthermore *Pms2*$^{-/-}$ mice have a level of MSI at mononucleotide repeats that is 2-3 fold lower than in *Mlh1*$^{-/-}$ mice (Yao *et al.*, 1999). This suggests that MLH1 can act in the repair of a subset of insertion:deletion loops either alone or with another partner, perhaps MLH3. In addition, the data demonstrate that MLH1 and PMS2 deficiencies have subtle, but differing effects on mutation avoidance and this may in turn have dramatic effects on tumour development. This might explain the different tumour spectrum between MLH1 and PMS2 deficient mice (Yao *et al.*, 1999).

1.5.4 Somatic inactivation of the MMR genes

Patients with HNPCC have one inherited MMR gene mutation and the remaining wild type allele is inactivated somatically during tumourigenesis, thus disrupting MMR activity. In HNPCC patients, inactivation of the second allele commonly arises

by somatic mutation (Leach *et al.*, 1993; Nicolaides *et al.*, 1994; Papadopoulos *et al.*, 1994). In kindreds with germline *MLH1* mutations, deletion of the wild type allele has also been observed, a mechanism know as loss of heterozygosity (LOH) (Hemminki *et al.*, 1994; Wheeler *et al.*, 2000). Evidence for LOH occurring at the *MSH2* locus has been examined using the closely linked marker, D2S123 but it appears that this locus is not as susceptible to deletion (Aaltonen *et al.*, 1993; Wheeler *et al.*, 1999). In sporadic CRCs with MSI, the inactivation of both MMR alleles must occur somatically and mutational inactivation of *MLH1*, *MSH2*, *MSH3*, *MSH6* and *PMS2* has been detected (Borresen *et al.*, 1995; Liu *et al.*, 1995b; Ma *et al.*, 2000; Mai *et al.*, 1999; Malkhosyan *et al.*, 1996). However, in contrast to the situation in familial tumours, LOH is not observed at either *MSH2* or *MLH1*, suggesting this may represent a mechanism of gene inactivation specific to HNPCC related cancers. (Wheeler *et al.*, 1999). There is also disparity between the frequency with which the MMR genes are mutated in familial and sporadic MSI^+ tumours. While *MLH1* and *MSH2* mutations predominate in HNPCC, inactivation of *MLH1* accounts for the majority of sporadic cases (Cunningham *et al.*, 1998; Herman *et al.*, 1998; Cunningham *et al.*, 2001). This is explained by the relative contribution from epigenetic inactivation of *MLH1,* in familial and sporadic CRCs.

It has been established that epigenetic mechanisms contribute to gene inactivation and in particular, silencing of gene expression by DNA methylation has been observed at *MLH1* in many MSI^+ CRCs (Cunningham *et al.*, 1998; Herman *et al.*, 1998; Jones and Laird, 1999; Lynch and de la Chapelle, 1999; Wheeler *et al.*, 2000). CpG islands are areas of about 1kb, commonly found near the promoters of widely expressed genes (Bird, 1986; Cross and Bird, 1995). They are usually unmethylated thus allowing gene expression, but de novo methylation of these regions, results in the silencing of the associated gene (Boyes and Bird, 1992; Cross, 1995). Hypermethylation of the 5'CpG island of the *MLH1* promoter has been detected in the majority of sporadic primary CRCs with MSI and this is generally associated with loss of MLH1 protein expression (Cunningham *et al.*, 1998; Herman *et al.*, 1998; Wheeler *et al.*, 1999; Cunningham *et al.*, 2001).

Wheeler *et al* addressed the relative contribution of different mechanisms of MMR gene inactivation by analysing sporadic MSI^+ CRC cell lines for mutations,

promoter methylation and LOH (Wheeler *et al.*, 1999). Mutation of *MLH1* together with hypermethlation of the promoter region was found to account for the majority of sporadic MSI^+ tumours (Wheeler *et al.*, 1999) and hypermethylation of the *MLH1* promoter represents the most significant mutational difference between sporadic and familial MSI^+ CRCs (Wheeler *et al.*, 2000). Although methylation of the *MLH1* promoter has been detected in some HNPCC CRCs the overall contribution is believed to be small (Cunningham *et al.*, 1998; Wheeler *et al.*, 2000).

1.6 The Mutator Phenotype

1.6.1 Microsatellite instability

Gross genetic alterations such as LOH and chromosomal gains in anueploid CRCs are well documented and have demonstrated that cancer cells contain many genetic changes in comparison to normal tissue (Peinado *et al.*, 1992). However, the link between MMR gene mutations and widespread instability at simple sequence repeats (SSRs) furthered the notion that a high mutation rate makes a significant and necessary contribution to the development of cancer (Loeb, 1994).

Microsatellites are tandem repeats of simple sequences that occur ubiquitously throughout the human genome. These SSRs occur in both coding and non-coding sequence, although they are most common in non-coding DNA (Toth *et al.*, 2000). SSRs comprise ~10-50 copies of 1-6bp motifs that can occur as perfect tandem repeats, imperfect (interrupted) repeats or as a combination of the two types (Weber, 1990). Microsatellites are often highly polymorphic in terms of length but are normally transmitted through mitosis and meiosis without alteration (Weissenbach *et al.*, 1992; Weber and Wong, 1993) Although no specific function has been demonstrated for non-coding microsatellite sequences, they do not necessarily represent "junk" DNA. Non-coding repetitive tracts have been suggested to be binding sites for DNA topoisomerase (Spitzner *et al.*, 1990). In addition, there is some evidence that the length of the SSR within promoter regions may influence transcriptional activity (Kashi *et al.*, 1997).

MSI was initially described in independent publications from three groups that appeared almost simultaneously in 1993 (Aaltonen *et al.*, 1993; Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). Each study used PCR based assays to compare loci between normal and tumour tissue and reported frequent alterations in the length of mononucleotide and dinucleotide repeats in tumour DNA (Aaltonen *et al.*, 1993; Ionov *et al.*, 1993; Thibodeau *et al.*, 1993). These alterations varied in nature and were observed as substantial changes in repeat length (often heterogeneous in nature), or as minor changes (typically 2bp) (Thibodeau *et al.*, 1993). In addition, the number of microsatellite loci that displayed length alterations varied from tumour to tumour. While changes were observed at a single locus in some tumours, alterations at many were detected in others (Thibodeau *et al.*, 1993). MSI was demonstrated to occur in most familial CRCs and 13% of sporadic cases, indicating the phenotype is a significant factor in CRCs and is characteristic of tumours from HNPCC families (Aaltonen *et al.*, 1993). It was recognised that although many of the changes may be silent, some may occur within coding regions and thus result in the significant alteration of gene products (Aaltonen *et al.*, 1993). The presence of generalised deletions at poly(A) and (CA)n repeats also correlated with a number of clinical features. The presence of MSI is negatively associated with mutations in *p53* and *C-K-RAS* but positively associated with poorly differentiated right-sided tumours (Ionov *et al.*, 1993).

The link between MSI and MMR defects was made by Strand *et al* who demonstrated that inactivating mutations in the yeast MMR genes, *MSH2*, *MLH1* and *PMS1* lead to a 100 to 700 fold increase in repeat tract instability (Strand *et al.*, 1993). It was proposed that a high rate of DNA polymerase slippage occurs in vivo at templates that contain simple repeats. The resulting alterations are usually corrected by MMR, but in the majority of HNPCC CRCs and a subset of sporadic tumours, defects in MMR means these errors remain un-repaired (Strand *et al.*, 1993). Parsons *et al* subsequently made the association in humans (Parsons *et al.*, 1993).

These studies dramatically contributed to the elucidation of the genetic mechanisms that cause CRC in HNPCC families and highlighted a role for an elevated mutation rate in tumourigenesis.

1.6.2 Classification of MSI status

Since almost all CRCs harbouring defects in MMR genes display MSI, establishing tumour MSI status has become an important tool in molecular pathology (Liu *et al.*, 1996).

Around 90% of CRCs from HNPCC patient's display MSI and thus, the assessment of MSI status is an important aspect of HNPCC screening (Aaltonen *et al.*, 1993). Exclusion criteria are applied that subsequently direct strategies to further screen the MMR genes for germline alterations (Rodriguez-Bigas *et al.*, 1997). Since up to 20% of sporadic tumours also display MSI, assessment of tumour MSI status is of great value in non-familial cases (Aaltonen *et al.*, 1993; Cunningham *et al.*, 2001). In a study of CRCs from non-HNPCC patients under 35 years of age, the majority (58%) exhibited a mutator phenotype and the underlying MMR gene mutations were also identified in a proportion of these patients (Liu *et al.*, 1995a). This suggests that an MSI pathway of tumourigenesis is of particular significance in certain subsets of sporadic CRCs especially in early onset cases.

MSI and MSS appear to characterise two different pathways of carcinogenesis and the determination of MSI status can have relevance to predicting the clinical course of a particular tumour. It is well documented that patients with MSI^+ tumours have a better prognosis than those whose tumours are MSS (Lothe *et al.*, 1993; Thibodeau *et al.*, 1993). Although, more recent evidence suggests that this may not always be the case (Farrington *et al* submitted). The fact that defects in MMR may result in resistance to DNA damaging agents has been discussed. Since MSI is indicative of MMR defects, identification of instability within a tumour may be predictive of its responsiveness to chemotherapy (Fink *et al.*, 1998).

Accurate determination of MSI status is critical in order to implement effective screening strategies and to make clinical and molecular predictions with confidence. However, it has been demonstrated that some microsatellite markers are more efficient than others in detecting tumours with MSI (Dietmaier *et al.*, 1997; Thibodeau *et al.*, 1998; Frazier *et al.*, 1999). While some markers are very unstable in MSI tumours, others show low levels of mutability. It has been suggested that the likelihood of a microsatellite being susceptible to instability relates to the inherent

mutation rate at that locus (Thibodeau *et al.*, 1998). This is supported by the observation of variation in mutation rates at di- tri- and pentameric markers within CEPH families (Weber and Wong, 1993). Variation in marker susceptibility to mutation can potentially affect the number of tumours that are scored positive for MSI. To address this problem, a number of studies have sought to identify a panel of markers that are the most efficient for detecting instability (Dietmaier *et al.*, 1997; Hoang *et al.*, 1997; Frazier *et al.*, 1999). At the National Cancer Institute (NCI) workshop in 1997, five markers were recommended for use as part of international guidelines for the evaluation of MSI in CRC (Boland *et al.*, 1998). This panel comprises, two mononucleotide markers, BAT25 and BAT26, and three dinucleotide markers D5S346, D2S123 and D17S250. A set of alternative loci was also suggested as part of an expanded panel and these include the poly(A) marker BAT40 (Boland *et al.*, 1998). Little is understood about factors independent of tumourigenesis that may influence the susceptibility of these markers to mutation in the absence of MMR. Elucidation of such factors is clearly of importance and further analysis is necessary, to fully characterise the nature of alterations that take place at these microsatellite loci in both tumour and normal cells (Boland *et al.*, 1998).

There is evidence that tumours themselves show differences in susceptibility to MSI and that MSI^+ actually represents two classes of tumours with distinct mutator phenotypes and different molecular aetiologies (Dietmaier *et al.*, 1997; Thibodeau *et al.*, 1998). Thibodeau *et al* noted two categories of tumours that either showed instability at a single repeat locus or showed instability at many (Thibodeau *et al.*, 1993). These two distinct groups have now been classified as tumours with high levels of MSI (MSI-H) or those with low MSI frequency (MSI-L) (Boland *et al.*, 1998). The international guidelines recommend that a tumour is classified as MSI-H if instability is detected at ≥2 of the recommended panel of five markers (Boland *et al.*, 1998). If one marker displays instability the tumour is classified as MSI-L and if all the markers are stable the tumour is MSS. However, the use of just five markers can result in cases of mis-classification between MSI-L and MSI-H (Frazier *et al.*, 1999). A more reliable analysis can be made by extending the number of loci used and classifying MSI-H tumours as those shifting at ≥ 30-40%, MSI-L as those

shifting at <30-40% and MSS as those displaying no microsatellite alterations (Boland *et al.*, 1998).

The sub-classification of MSI tumours into MSI-L and MSI-H is particularly useful as it can assist in screening strategies for MMR mutations. While most tumours with an MSI-H phenotype have defects in *MLH1* and *MSH2* (Liu *et al.*, 1996), mutations in *MSH6* appear to contribute to a substantial proportion of patients with MSI-L tumours and result in a more atypical HNPCC phenotype (Wu *et al.*, 1999b). In addition, the clinical phenotype of MSI-L tumours seems more similar to MSS tumours. For example, the predilection for right-sided tumours in MSI-H cases is not observed in MSI-L. Such observations suggest that the level of instability directly influences the clinical development of CRC (Thibodeau *et al.*, 1998).

Further characterisation of factors that can affect stability of microsatellite markers will improve the robustness of using these tools in the classification of MSI in CRCs. In turn this will allow more effective conclusions to be made with regard to the associations of relative levels of mutation within tumours and the association with clinical phenotype.

1.6.2 Role of MSI in tumourigenesis

The subtle changes in the DNA sequence caused by MSI, provide an alternative mutational mechanism to the cytogenetically visible changes such as chromosomal losses, gains and translocations that are well documented in the majority of solid tumours (Peinado *et al.*, 1992; Lengauer *et al.*, 1998). Tumours that show instability at the sequence level (MSI^+) are termed MIN, whereas those that are unstable at the chromosomal level (MSS) are known as CIN. These two types of instabilities are rarely found to co-exist in tumours and thus it appears that one form of instability is sufficient to promote tumourigenesis. The exact mechanisms leading to a CIN phenotype are largely unclear, however the relationship between MMR and MSI is well established and is clearly associated with the development of MIN tumours. Therefore, MIN tumours are of particular use in the investigating the function of increased genetic instability in tumourigenesis.

Since the discovery of MSI in CRCs, a number of studies have addressed the timing of this event and the role that an abnormally elevated mutation rate plays in the development of cancer (Aaltonen *et al.*, 1994; Shibata *et al.*, 1994). Shibata *et al* examined patterns of MSI at different regions of colorectal carcinomas and adenomas and detected MSI in colorectal adenomas, suggesting that genomic instability is a very early event (Shibata *et al.*, 1994). That MSI arises early in tumour development is supported by the detection of genetic alterations in aberrant crypt foci (ACF), which represent morphologically abnormal crypts, believed to be the precursors of adenomas (Augenlicht *et al.*, 1996; Heinen *et al.*, 1996). Since alterations in SSRs are not generally observed in normal tissue the activation of MSI appears strongly associated with histological transformation and is implicated as having a critical role in tumour initiation (Aaltonen *et al.*, 1994; Shibata *et al.*, 1994). There are some exceptions in which MSI has been detected in non-cancerous tissue but these cases are generally associated with constitutional defects in MMR (Parsons *et al.*, 1995a; Hackman *et al.*, 1997; Narayanan *et al.*, 1997; De Rosa *et al.*, 2000).

MSI has been demonstrated to display different degrees of instability within different regions of a given tumour and these regions comprise areas showing distinct clonality. In addition more advanced cancers display greater instability compared to early adenomas (Shibata *et al.*, 1994). These observations provide evidence that genomic instability is a dynamic process, which persists throughout tumour development and likely contributes to both tumour progression and initiation (Shibata *et al.*, 1994).

A model of tumourigenesis driven by genomic instability has been proposed (Figure 1.4). In this model, tumours initiate as a result of inactivation of a single MMR gene. Consequently, mutations accumulate in the offspring of this cell and waves of clonal expansion give rise to daughter cells that have a growth advantage typical of cancer (Nowell, 1976; Fearon and Vogelstein, 1990; Cahill *et al.*, 1999; Loeb, 2001). The mutational targets of this multistep progression are oncogenes and tumour suppressor genes that when mutated, directly affect the rate of cell growth or cell death. The increase in mutation rate in MIN tumours provides variation for selection to act upon. However, it is reasonable to predict that the intrinsic level of instability may be critical (Cahill *et al.*, 1999). If the mutation rate is too low, cells

will be unlikely to contain an alteration that provides them with a selective advantage. Conversely, if the mutation rate is too high and the cell becomes overloaded with damage, programmed cell death will ensue (Cahill *et al.*, 1999). Advocates of the mutator phenotype propose that MSI is one of the earliest events that drives tumour progression and furthermore, is necessary to generate the level of mutational alterations on which selection may then act (Loeb, 2001).

The molecular pathogenesis of MIN tumours remains to be fully elucidated although some causative genes have been isolated. Identification and functional understanding of the mutations that accumulate in coding sequences in MMR defective tumours, will be crucial to understanding mechanisms by which genomic instability contributes to tumourigenesis.

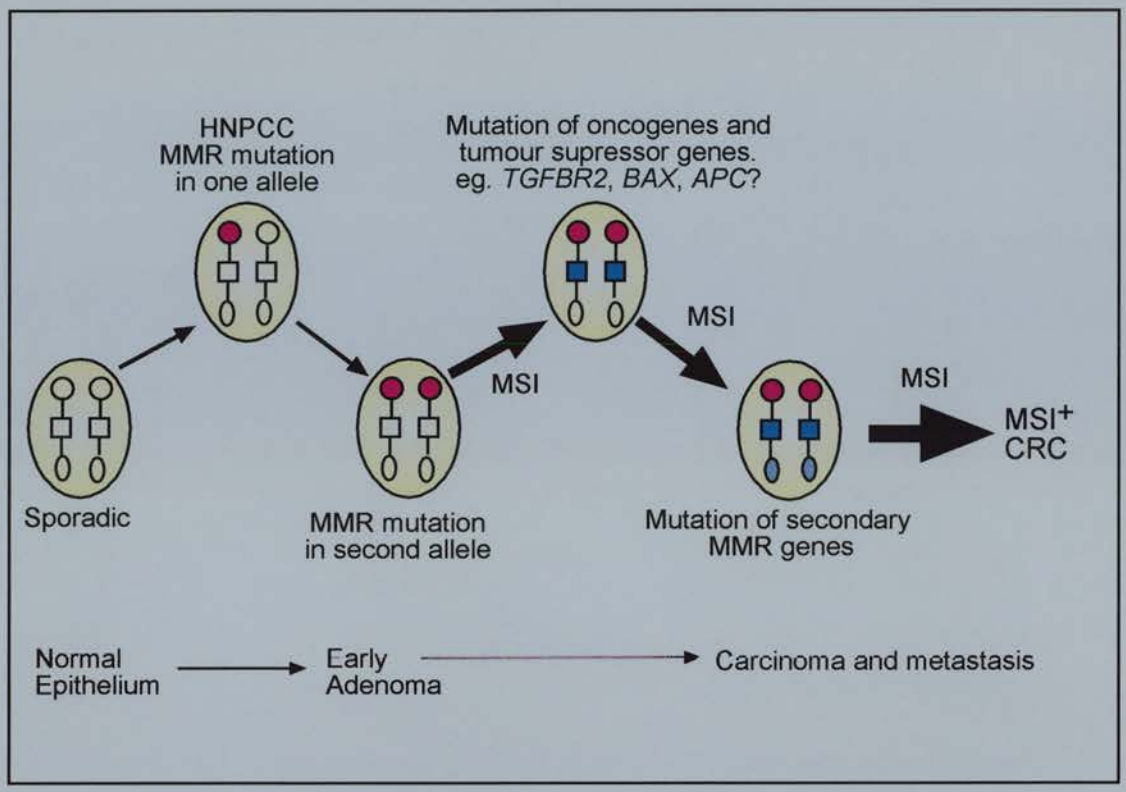


Figure 1.4 Generalised model of the mutator pathway of colorectal cancer. The thickness of the arrows is intended to convey the gradual increase in the probability of occurance of mutation in genes required for tumour progression. The mutational events indicated are not definitive but represent probable targets as suggested from the literiture. It is highly likely that many more mutations than indicated here, contribute to this pathway of tumour development. Gene alterations are different to those recognised in the initial pathway proposed by Vogelstein and Fearon, 1991, as is the speed with which changes arise once tumourigenesis is initiated (Figure 1.1). Adapted from Perucho, 1996; Eshleman and Markowitz, 1996 and Boland, 2000.

1.7 Molecular Consequences of Mismatch Repair Defects

1.7.1 Coding sequence mutations in MMR deficient tumours

Microsatellite repeats are clearly hotspots for mutation in MMR deficient tumours. Repeat regions in coding sequence have been identified and their mutational status evaluated in MSI^+ CRCs, in order to elucidate critical oncogenes and tumour suppressor genes involved in the mutator pathway of CRC (Parsons *et al.*, 1995b; Duval *et al.*, 2001). Repetitive tracts, particularly poly(A/T) mononucleotide repeats and CA:GT dinucleotides are common in human coding sequence although, they tend to be much shorter than in non-coding sequences (Toth *et al.*, 2000) (Taylor pers. comm.). However, due to the high background of genetic instability that characterises MSI^+ tumours, it is often difficult to establish which alterations play a crucial role in the initiation and progression of MSI^+ CRCs. An elevated frequency of mutations within a gene in MSI^+ tumours does not necessarily imply a causal role for the mutation in CRC development.

To distinguish between mutations occurring in target genes and those that are merely bystanders, a set of five criteria has been established (Boland *et al.*, 1998). In addition to (1) having a high mutational frequency in MSI^+ CRCs, genuine targets involved in tumourigenesis should (2) also demonstrate biallelic inactivation. It is reasonable to suggest that only when both alleles of the putative target are inactivated, is it highly likely that protein function will be completely abrogated. (3) A role for the candidate gene within a tumour suppressor pathway should be apparent, indicating a potential mechanism by which inactivation of the gene provides the cell with a growth advantage. (4) The occurrence of alterations within the same gene or same functional pathway in MSS tumours may also be expected if the target is critical for tumour suppression. Finally, (5) functional studies carried out in vitro and in vivo, should be consistent with a role for the protein, in growth suppression (Boland *et al.*, 1998). This set of criteria is not without limitations. By suggesting that a role for the gene in a tumour suppressor pathway should be implicated (point 3), this assumes that all important tumour suppressor pathways are known. Clearly this is not the case. In addition, genes inactivated in MSI^+ tumours

may not also be inactivated in MSS tumours (point 4) if the genetic pathways involved in the development these tumours is different and there has been several studies to suggest this may be the case (Olschwang *et al.*, 1997; Perucho, 1999).

Models for the progression of CRC have been presented as an ordered sequence of mutations in oncogenes and tumour suppressor genes, each associated with a defined step in tumourigenesis (Figure 1.1) (Vogelstein, 1988; Fearon, 1990). However, a mutator phenotype may not be in accord with a set order of mutation during tumour evolution, due to the large numbers of apparently different and random mutations that occur and confer potential selective advantage (Loeb, 2001). Nonetheless, a number of genes have been implicated as critical targets in a mutator pathway of CRC development. It may follow that certain gene mutations may arise more readily if particular sequences are inherently susceptible to mutation consequent of MMR defects.

1.7.2 *TGFBR2* mutations in MSI$^+$ CRCs

The transforming growth factor beta type 2 receptor (TGFBR2) has been implicated as having a major role in the development of MSI$^+$ CRCs (Markowitz, 2000). There are several lines of evidence that suggest the *TGFBR2* gene is genuinely inactivated in CRCs with MMR defects and that this inactivation directly contributes to tumourigenesis.

The transforming growth factor beta, (TGFB) signalling pathway results in potent antiproliferative responses (Wang *et al.*, 1995). Loss of this negative regulatory mechanism causes excess cell growth, suggesting that the *TGFBR2* gene behaves as a tumour suppressor. Antiproliferative responses triggered by TGFB include, induction of growth arrest in late G1 phase of the cell cycle (Alexandrow and Moses, 1995) and induction of apoptosis (Wang *et al.*, 1995) and these activities have been demonstrated in non-transformed colon epithelial cells (Wang *et al.*, 1995). That TGFBR2 is functionally a colon cancer suppressor gene has been demonstrated by the restoration of wild type function by addition of wild type *TGFBR2* into *TGFBR2* mutant CRC cell lines (Grady *et al.*, 1999).

TGFB signals, by contacting the two distantly related transmembrane serine/threonine kinases called receptors I and II (TGFBR1 and TGFBR2) (Figure 1.5) (Lin *et al.*, 1992; Wrana *et al.*, 1994). TGFB binds directly to TGFBR2, a constitutively active kinase. TGFB bound to TGFBR2 is recognised by TGFBR1 and this second receptor is then recruited to form a heteromeric receptor complex. On formation of this complex, TGFBR1 is phosphorylated by TGFBR2 (Wrana *et al.*, 1994; Massague, 1996; Markowitz, 2000). The resultant activated receptor, signals to the nucleus via members of a family of signal transduction molecules called SMADs, which are related to the *Drosophila* gene, Mothers against dpp (Mad) and the *C. elegans* sma homologues (Zhang *et al.*, 1996; Nakao *et al.*, 1997; Wrana and Pawson, 1997; Attisano and Wrana, 1998; Kretzschmar and Massague, 1998). Specific receptor activated SMADs, (SMAD2 and SMAD3) are recognised by the activated TGFB receptor complex and are phosphorylated at carboxy-terminal serine residues (Eppert *et al.*, 1996; Nakao *et al.*, 1997; Wrana and Pawson, 1997). These molecules are then able to associate with a co-SMAD, (SMAD4) and translocate to the nucleus where Smad2-Smad4 complexes have been demonstrated to regulate TGFB signalling responses by specifically interacting with DNA-binding proteins such as the transcription factor FAST-1 (Figure 1.5) (Kretzschmar and Massague, 1998; Nakao *et al.*, 1997; Wrana and Pawson, 1997). The regulatory activities of TGFB include inhibiting expression and activation of cyclins, inhibiting expression of cyclin dependent kinases, inducing various cyclin-dependent kinase inhibitors such as p15, p27 and p21, inhibiting phosphorylation of Rb and inhibiting induction of c-myc (Moses *et al.*, 1990; Ewen *et al.*, 1993; Hannon and Beach, 1994; Markowitz, 2000). Inactivation of any of these targets could explain why human cancer cells show resistance to TGFB mediated growth arrest.

Inactivating mutations have been identified in *TGFBR2* in over 90% of MSI^+ CRCs (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b). These mutations are principally frameshift mutations that insert or delete one or two adenine bases within a 10bp polyadenine repeat $poly(A)_{10}$ that is present in exon 3 of the *TGFBR2* gene (Markowitz *et al.*, 1995; Lu *et al.*, 1996). The mutations observed introduce stop codons in the *TGFBR2* transcript. These result in a truncated protein that lacks the wild type cytoplasmic domain and is therefore functionally inactive.

Parsons *et al*, analysed 111 MSI^+ primary CRC tumours and cell lines specifically for mutations within the $poly(A)_{10}$ tract of *TGFBR2* and identified frameshift mutations in 100 (90%). In most cases, biallelic inactivation of *TGFBR2* was observed (Parsons *et al.*, 1995b). Since the mutations were found in tumour DNA, i.e.) not in corresponding normal DNA, these mutations must have been somatic. Furthermore, although poly(A) tracts are generally unstable in MSI^+ tumours (Ionov *et al.*, 1993; Thibodeau *et al.*, 1993), the $poly(A)_{10}$ tract of *TGFBR2* appeared significantly more unstable than other poly(A) tracts of the same size (Parsons *et al.*, 1995b).

Further studies have addressed the timing of the *TGFBR2* $poly(A)_{10}$ mutations observed in MSI^+ CRCs. Those genetic changes critical for carcinogenesis, are likely to be shared by all cells in an expanding tumour and thus be detectable at all sites (Abdel-Rahman *et al.*, 1999). Consistent with this notion, *TGFBR2* $poly(A)_{10}$ mutations were identified at all tumour sites sampled in 9 MSI^+ CRCs (Abdel-Rahman *et al.*, 1999). Furthermore, *TGFBR2* $poly(A)_{10}$ mutations have also been detected in early colorectal lesions (Lu *et al.*, 1996; Akiyama *et al.*, 1997a; Grady *et al.*, 1998). Akiyama *et al* revealed altered *TGFBR2* $poly(A)_{10}$ repeats in all 5 adenomas sampled with moderate atypia (Akiyama *et al.*, 1997a). It was suggested that mutation of *TGFBR2* occurs early in tumourigenesis, functioning to allow MMR defective cells to proceed to adenoma formation (Akiyama *et al.*, 1997a). Grady *et al* also detected *TGFBR2* $poly(A)_{10}$ mutations in 75% of late stage MSI^+ adenomas, not significantly different to the frequency observed in carcinomas. This observation suggests that the inactivation of TGFBR2 is tightly correlated with the progression of late stage adenomas to cancer (Grady *et al.*, 1998). Mutations at this region have not been observed in normal cells, even those with constitutive defects in MMR, although few studies have analysed this locus in non-cancerous tissue (De Rosa *et al.*, 2000). Therefore the nature of the mutations (predominantly frameshifts) at this microsatellite locus and their timing is consistent with them having arisen early in tumourigenesis, but subsequent to the loss of MMR (Grady *et al.*, 1998; Abdel-Rahman *et al.*, 1999).

Further evidence that mutation of *TGFBR2* is a key event in tumours defective in MMR, is suggested by the observation of inactivating point mutations in *TGFBR2* in

MSS CRCs (Grady *et al.*, 1999). Inactivation of other components of the TGFBR2 pathway has also been identified in a number of MSS cancers. *SMAD4*, also known as *DPC4*, was originally isolated as a tumour suppressor gene on 18q21 that was either deleted or mutated in large numbers of human pancreatic carcinomas (Hahn *et al.*, 1996). *SMAD2* is also located at 18q21 and is mutated in MSS CRCs (Eppert *et al.*, 1996). These observations further suggest that inactivation of the TGFB signalling pathway is an important step in the development of CRC.

Taken together the evidence suggests that mutations arise frequently in the poly$(A)_{10}$ tract of the *TGFBR2* gene and are critical to MSI^+ associated tumourigenesis. The mutational inactivation of TGFBR2 may be involved in initiating events of CRCs with MMR defects and loss of TGFBR2 likely confers a strong selective advantage to the developing tumour.

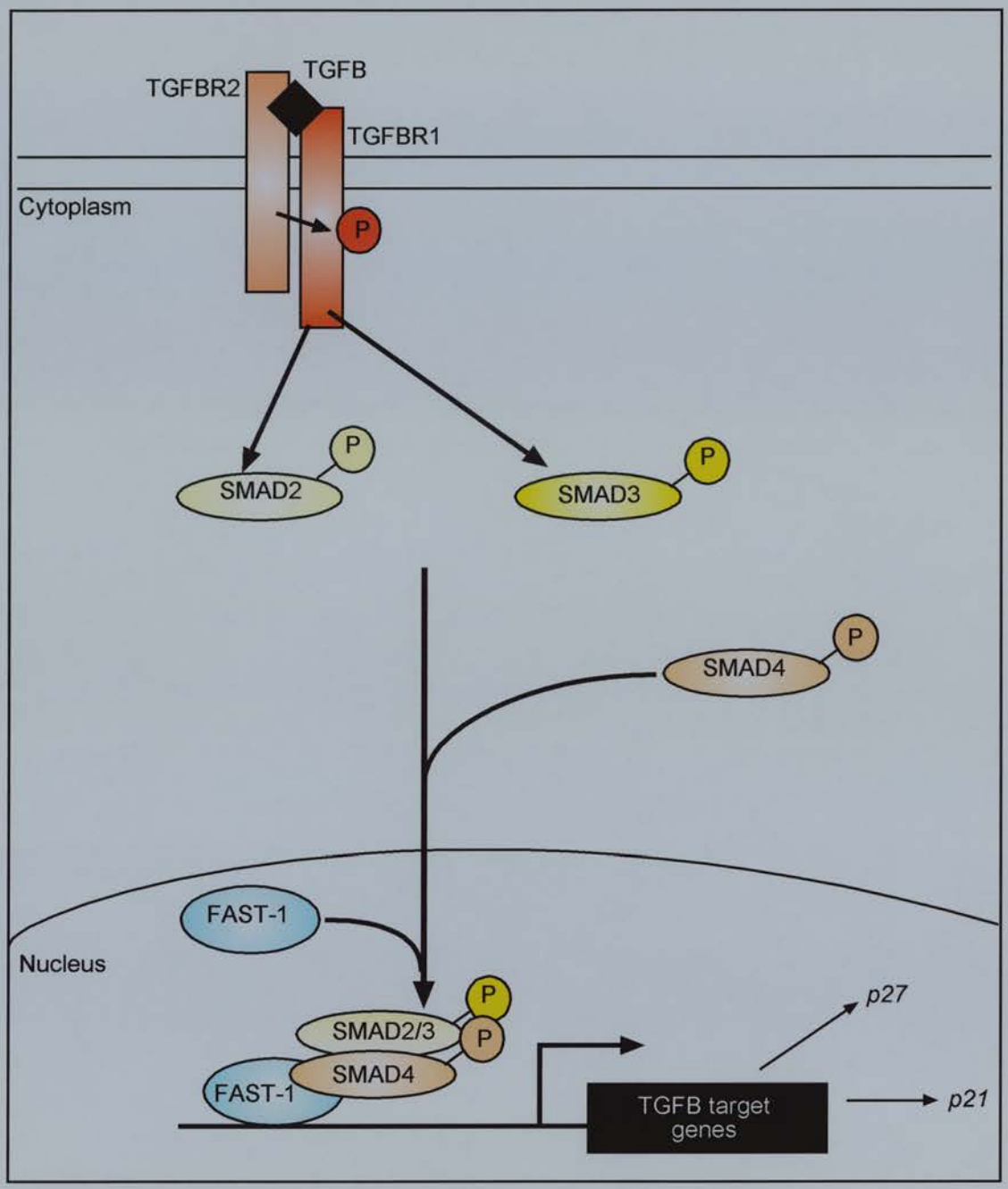


Figure 1.5 General model of TGFB/SMAD signalling. TGFB signalling induces phosphorylation of TGFBR1 by TGFBR2 leading to the transient association with specific receptor activatable SMADS (2 and 3). On phosphorylation they associate with a co-SMAD (SMAD4) and move to the nucleus. Formation of a complex with DNA binding transcription factors such as FAST-1 leads to stimulation of target gene expression. Adapted from Nakao *et al.*, 1997; Kretschmar and Massague, 1998, and Markowitz, 2000.

1.7.3 *BAX* mutations in MSI$^+$ CRCs

Around 50% of MSI$^+$ CRCs contain frameshift mutations of a poly(G)$_8$ tract in exon 3 of the pro-apoptotic *BAX* gene and in some cases bi-allelic inactivation has been detected (Rampino *et al.*, 1997; Ouyang *et al.*, 1998). Although the overall frequency of these mutations is not as high as mutations of the poly(A)$_{10}$ tract in *TGFBR2,* the prevalence of mutations observed in the repetitive tract of *BAX* are still in excess of those observed in sequences of similar size in MSI$^+$ tumours (Rampino *et al.*, 1997; Ouyang *et al.*, 1998; Zhang *et al.*, 2001). In addition, there is evidence that mutational inactivation of the *BAX* gene confers a selective advantage during clonal evolution of the tumour (Ionov *et al.*, 2000).

BAX is a member of an extended family of proteins that are activated by cleavage. This cleavage may be carried out by caspases, by inhibition of protein kinases and/or activation of phosphatases and by an increase in cellular pH (Pawlowski and Kraft, 2000). BAX molecules reside in the cytoplasm and respond to apoptotic stimuli by migration to the mitochondria (Figure 1.6). BAX contains a hydrophobic membrane anchor at its C-termini that is necessary for mitochondrial insertion (Goping *et al.*, 1998). Attachment to the mitochondria allows BAX to directly trigger the release of cytochrome C, by inducing the formation of ion-permeable pores that disrupt the mitochondrial membrane barrier (Jurgensmeier *et al.*, 1998; Marzo *et al.*, 1998). Upon entering the cytosol, cytochrome C promotes the assembly of a multiprotein complex containing APAF-1, that induces proteolytic processing and activation of cell death proteases, known as caspases. Bcl-2 is an antagonist of BAX that prevents the release of cytochrome C, and while BAX and Bcl-2 can function independently, their balance regulates the apoptotic response of the cell to the same apoptotic stimuli (Oltvai *et al.*, 1993; Pawlowski and Kraft, 2000). Thus, the function of BAX suggests that its inactivation may play a direct role in tumourigenesis since cells lacking a functional protein may have diminished capacity to trigger apoptosis upon receiving a death signal (Figure 1.6) (Oltvai *et al.*, 1993; Rampino *et al.*, 1997; Yin *et al.*, 1997).

The p53 protein plays a role in triggering apoptosis in response to DNA damage and has been demonstrated to transactivate BAX (Miyashita and Reed, 1995;

Thornborrow and Manfredi, 1999). Therefore, the finding of an appreciable frequency of *BAX* mutations in MSI^+ CRCs may explain why in contrast to MSS tumours, those of the mutator pathway do not contain p53 mutations (Fearon and Vogelstein, 1990; Ionov *et al.*, 1993; Kim *et al.*, 1994; Konishi *et al.*, 1996; Olschwang *et al.*, 1997; Yagi *et al.*, 1998). In an MMR defective cell, mutations at the *BAX* poly$(G)_8$ repeat are more likely to occur than other frameshift or missense mutations in p53. Inactivation of BAX thus eliminates the selective pressure for p53 mutations in MSI^+ CRCs, since an apoptotic signalling molecule downstream of p53 has already been lost (Rampino *et al.*, 1997; Yagi *et al.*, 1998).

While *TGFBR2* mutations appear to occur relatively early in tumourigenesis, the lack of clonality of *BAX* poly$(G)_8$ mutations in some MSI^+ carcinomas, indicates that these changes are unlikely to be such an early event (Abdel-Rahman *et al.*, 1999). *BAX* poly$(G)_8$ mutations have been observed in only 15% of adenomas, significantly less than observed in adenocarcinomas (54%), suggesting that these mutations show a high mutation rate in the transition to carcinomas (Yagi *et al.*, 1998). It has also been suggested that *BAX* mutations might not exert a critical role in tumour initiation, but may aid tumour progression (Abdel-Rahman *et al.*, 1999).

The high frequency of mutations of the repeat tract of the *BAX* gene in MSI^+ tumours along with the functional evidence that inactivation can confer a selective advantage during selection, suggests that the *BAX* poly$(G)_8$ tract is an important mutational target in MSI^+ CRCs. The difference in mutation frequency of the repeat regions in *TGFBR2* and *BAX* may reflect differences in inherent stability of the repeats in the presence of repair defects or differences in selectable advantage conferred by such mutations.

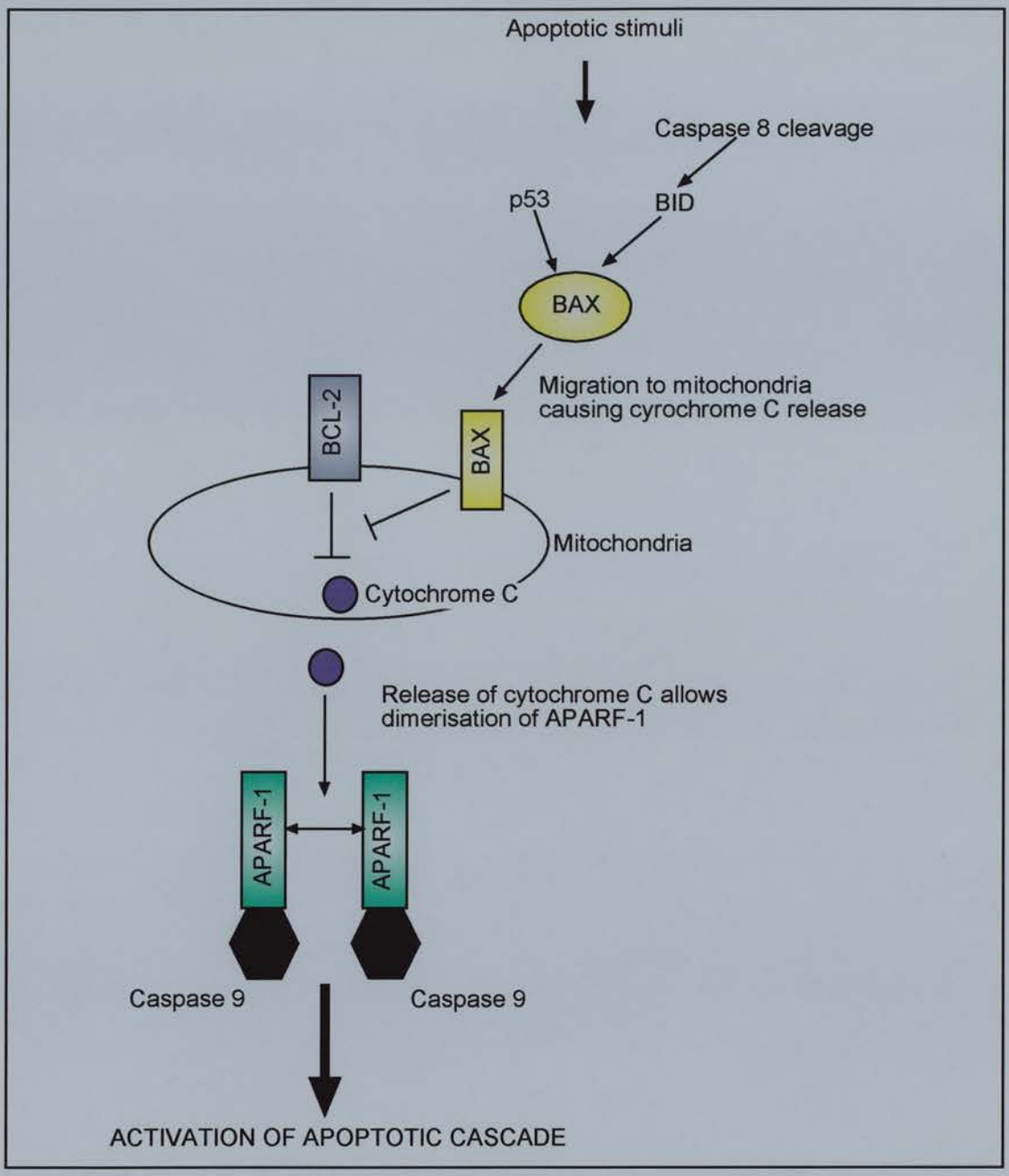


Figure 1.6 Cascade of the BAX mediated apoptotic response. Major componants involved in the pro-apoptic response by BAX are illustrated. p53 and BID represent examples of effector molecules stimulated by apoptitic stimuli. On release of cytochrome C from the mitochondria the two caspase 9 molecules are brought together by APARF-1 thus activating this initiator caspase. Activated caspase 9 induces cleavage of effector proteases, caspase 3 and 7 activating their protease activity. These are responsible for the controlled degredation process characteristic of apoptotic cell death including activation of the DNA cleavage nuclease (CAD). Adapted from Pawlowski and Kraft, 2000 and Macleod, 2000.

1.7.3 Inactivation of the β-catenin/TCF pathway in MSI$^+$ CRCs

The tumour suppresser gene *APC* was shown to cause the syndrome FAP by observing co-segregation of mutant *APC* alleles in affected kindreds (Groden *et al.*, 1991; Kinzler *et al.*, 1991; Nishisho *et al.*, 1991). Following this report, mutations of the *APC* gene were demonstrated to occur in around 80% of sporadic CRCs and adenomas, indicating that the *APC* gene plays a major role in the early development of colorectal neoplasms (Miyoshi *et al.*, 1992; Powell *et al.*, 1992). It has been suggested that APC acts as a gatekeeper of colonic epithelial cell proliferation and that inactivation of this gatekeeper is required for an overall increase in cellular proliferation in the majority of CRCs (Kinzler and Vogelstein, 1996). Although *p53* and *C-K-RAS* mutations are also frequently observed in sporadic CRCs, germline mutations in these genes do not pre-dispose to colon cancer, suggesting they cannot initiate tumourigenesis in a manner similar to inactivation of APC (Kinzler and Vogelstein, 1996).

Insights into APC function were gained from studies of the interaction between APC and β-catenin (Rubinfeld *et al.*, 1993; Su *et al.*, 1993) Interestingly, the majority of somatic mutations in the *APC* gene occur in the mutation cluster region, a 666bp region that contains two classes of β-catenin binding repeats (Figure 8.2). These comprise 3, 15 amino acid (aa) signature repeats that provide binding sites for β-catenin and seven 20aa repeat motifs that allow β-catenin binding and down regulation (Miyoshi *et al.*, 1992; Rubinfeld *et al.*, 1996; Fearnhead *et al.*, 2001). At least three of the 20aa repeats are necessary for partial down regulation of β-catenin by APC and four or more for a complete effect (Su *et al.*, 1993; Munemitsu *et al.*, 1995; Fearnhead *et al.*, 2001). In CRCs containing *APC* mutations, five or more of these repeats are typically deleted (Rubinfeld *et al.*, 1997). This suggests that abrogation of APC binding and regulation of β-catenin is selectively disrupted in colorectal tumours (Munemitsu *et al.*, 1995; Rubinfeld *et al.*, 1997).

β-catenin plays a role in cadherin mediated cell adhesion. Binding to cadherins is mutually exclusive to binding of APC, indicating that APC could modulate such adhesion as part of its gatekeeping role (Barth *et al.*, 1997). β-catenin and APC also

play major roles in the β-catenin/TCF signalling pathway (Figure 1.7) (Munemitsu *et al.*, 1995; Barth *et al.*, 1997; Morin *et al.*, 1997) and it is the disruption of this pathway that is thought to be of particular significance in CRC development. Abrogation of the β-catenin/TCF pathway leads to tumour formation most likely via the constitutive activation of oncogenic targets that include cyclin D1, c-myc and c-jun (Munemitsu *et al.*, 1995; Mann *et al.*, 1999; Tetsu and McCormick, 1999 Behrens, 2000). Understanding of this signalling pathway in humans, has been elucidated from studies of the Wingless (Wg) and Wnt pathways, in *Drosphila* and mouse respectively, where β-catenin has been established as the central signal transducer in this pathway (Behrens, 2000). The control of cytoplasmic β-catenin levels is an essential regulatory mechanism that prevents inappropriate Wnt signal transmission to the nucleus when Wnt factors are not present (Munemitsu *et al.*, 1995; Behrens, 2000). In the absence of Wnt signalling, a cytoplasmic multisubunit complex is assembled that targets β-catenin for degradation. Components of this complex include conductin/axin, a serine/threonine kinase called glycogen synthase3β (GSK3β) and APC (Rubinfeld *et al.*, 1996). Degradation of β-catenin is induced by the phosphorylation of specific serine/threonine residues in its N-terminus probably by GSK-3β. This leads to ubiquitination of β-catenin and subsequent proteosome degradation (Orford *et al.*, 1997). Degradation is restricted to the cytoplasmic pool of the β-catenin and does not involve protein bound to cadherins. In the presence of Wnt signals, cytoplasmic β-catenin is stabilised. When present at high levels in the cytoplasm, β-catenin binds to cytosolic T cell factor-4/lymphoid enhancer factor (TCF/LEF) proteins and the resulting complex is shuttled to the nucleus (Behrens *et al.*, 1996). As a consequence, the β-catenin/TCF complexes, activate transcription of specific target genes (Mann *et al.*, 1999).

The essential function of APC in the β-catenin regulatory complex has not been fully elucidated. However, the finding that protein products of mutant *APC* genes that are present in colorectal tumours, loose their ability to down-regulate transcriptional activation mediated by β-catenin and T-cell transcription factor, indicates that regulation of β-catenin is critical to APCs tumour suppressive effect (Morin *et al.*, 1997). Furthermore, in sporadic CRCs lacking *APC* mutations,

activating mutations in the β-catenin gene, *CTNNB1*, have been identified that alter functionally significant phosphorylation sites (Morin *et al.*, 1997). These *CTNNB1* mutations have been demonstrated to encode proteins that circumvent the tumour suppressive effect of APC (Morin *et al.*, 1997).

The role of the β-catenin/TCF signalling pathway in tumourigenesis of CRCs with MMR defects is unclear. Evidence from mice studies has suggested that the loss of APC is involved in HNPCC. For example, *Msh2*$^{-/-}$ mice show loss of detectable APC protein in their tumours (Reitmair *et al.*, 1996). In addition, when an APC mutation was bred into the *Mlh1*$^{-/-}$ mouse, the incidence of GI tumours increased 40-100 fold (Edelmann *et al.*, 1999). However, in humans the reported frequency of APC mutation in MSI^{+} CRCs differs markedly. A number of studies have demonstrated that *APC* mutations occur infrequently in MSI^{+} CRCs and that these tumours involve the abrogation of alternative genes and genetic pathways compared to MSS CRCs (Heinen *et al.*, 1995; Konishi *et al.*, 1996; Olschwang *et al.*, 1997; Salahshor *et al.*, 1999). Salahshor *et al* also screened MSI^{+} tumours for *CTNNB1* mutations since these can substitute *APC* mutations in MSS CRCs (Morin *et al.*, 1997; Salahshor *et al.*, 1999). However, in 22 MSI^{+} tumours no *CTNNB1* mutations were identified furthering the argument that activation of the β-catenin/TCF pathway is not critical for tumour development when MMR is defective (Salahshor *et al.*, 1999). In contrast, a number of reports have demonstrated that the frequency of *APC* mutations in MSI^{+} CRCs is similar to that of MSS CRCs (Huang *et al.*, 1996; Homfray *et al.*, 1998) and *CTNNB1* mutations have been detected in a number of MSI^{+} tumours lacking *APC* mutations (Sparks *et al.*, 1998; Mirabelli-Primdahl *et al.*, 1999; Miyaki *et al.*, 1999;). Abrogation of β-catenin has even been suggested to be an event specific to MSI^{+} CRCs (Mirabelli-Primdahl *et al.*, 1999).

Where *APC* mutations have been observed frequently in MSI^{+} CRCs, there is uncertainty as to whether they occurred prior to inactivation of MMR or as a consequence of these defects. This brings into question the role of an increased mutation rate being necessary for initiation and development of MSI^{+} CRCs (Huang *et al.*, 1996; Homfray *et al.*, 1998; Lamlum *et al.*, 2000). Although the *APC* gene does not contain a specific repetitive tract that may be particularly prone to mutation

consequent of repair defects, it does contain many sites of short repeated sequences, especially poly(A) mononucleotide repeats (Joslyn *et al.*, 1991). Huang *et al* found that although the frequency and distribution of mutations at the *APC* gene was similar between MSI^+ and MSS CRCs, the mechanism of mutation varied. Mutations in MSS tumours were predominantly point mutations, whereas in MSI^+ tumours the mutation spectrum was significantly biased towards frameshift mutations within poly(A) tracts, characteristic of defects in MMR (Ionov *et al.*, 1993; Strand *et al.*, 1993; Thibodeau *et al.*, 1993). In contrast, Homfray *et al* reported no difference in the mutational spectrum between MSI^+ and MSS CRCs (Homfray *et al.*, 1998). The same group also detected *APC* or β-catenin mutations in early adenomas from sporadic CRCs but no evidence of MSI (Lamlum *et al.*, 2000). These observations suggests that *APC* mutations arise before MMR mutations as a consequence of selection and are of greater significance in tumourigenesis than the elevated mutation rate that may occur later due to MMR defects (Homfray *et al.*, 1998; Lamlum *et al.*, 2000; Sieber *et al.*, 2000).

Mutations of repetitive elements within the coding sequence of other components of the β-catenin/TCF pathway in MSI^+ CRCs have been reported. There is some evidence that mutations in *TCF-4* and *AXIN2* (Conductin) may contribute to disruption of this pathway. Mutation of *TCF-4* appears almost exclusive to MSI^+ tumours and 1bp deletions of an $(A)_9$ repeat have been detected in 50% of MSI-H cell lines (Duval *et al.*, 1999). However the cell lines, in which these mutations were identified, also had either *APC* or *CTNNB1* mutations and thus the relative contribution from the *TCF-4* mutations is unclear (Duval *et al.*, 1999). *AXIN2*, the human homologue of mouse Conductin, was also mutated in 11 out of 45 MSI^+ CRCs (Liu *et al.*, 2000). Mutations were detected in one of the four mononucleotide repeats located in exon 7 and were demonstrated to activate β-catenin/TCF dependent transcription. *APC* and *CTNNB1* mutations were not found in tumours with *AXIN2* mutations suggesting that it was the *AXIN2* mutations that contributed to the development of CRC in these patients (Liu *et al.*, 2000).

Although activation of the β-catenin/TCF pathway is a critical event in the development of the majority of CRCs, its role in MSI^+ tumourigenesis remains

unclear. Conclusive evidence is required to define whether abrogation of this 'gatekeeper' pathway is necessary in MSI^+ tumourgenesis and whether inactivation of components of this signalling pathway can occur consequent of repair defects.

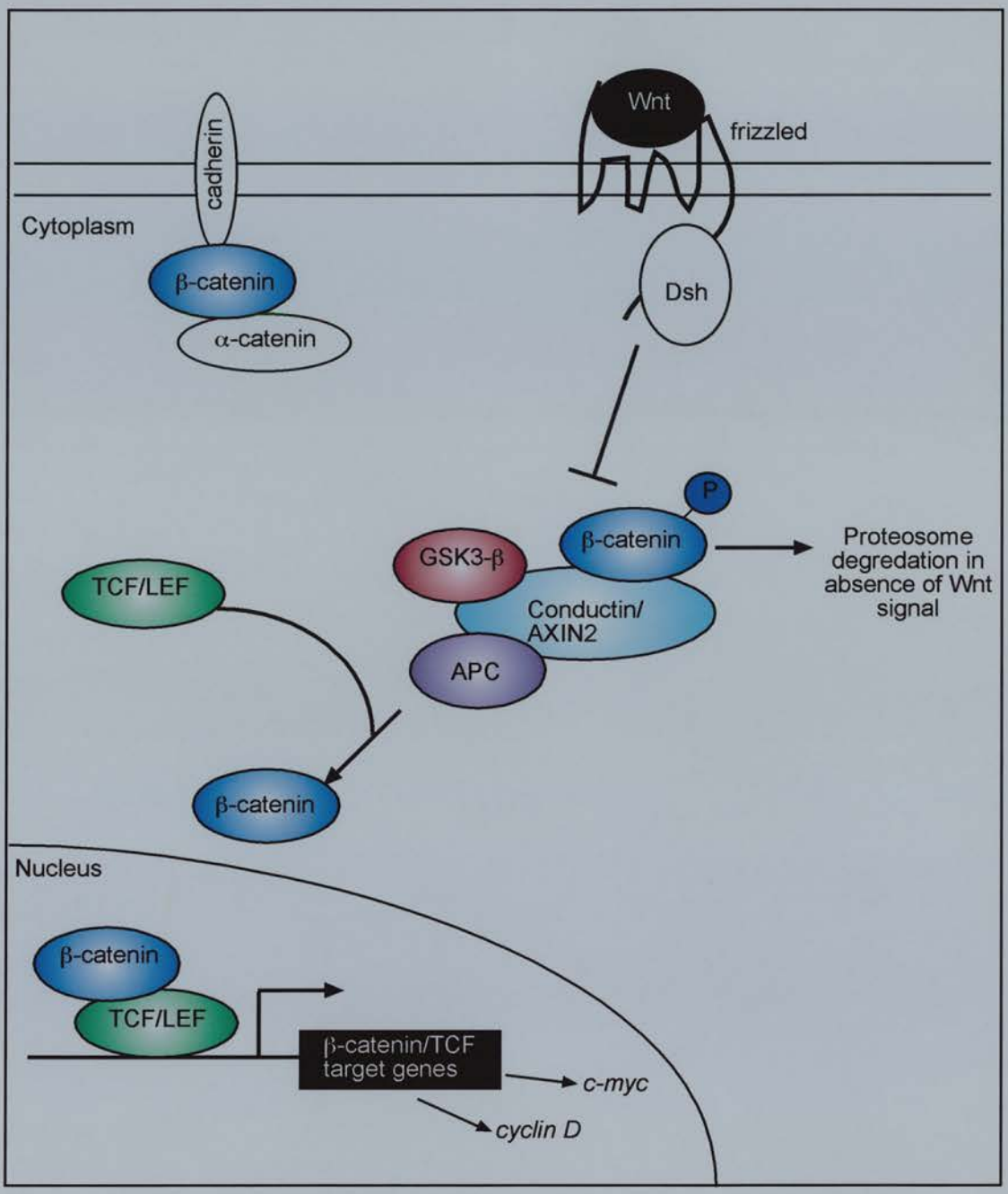


Figure 1.7 The β-catenin/TCF pathway. In the absence of a Wnt signal the phosphorylation of β-catenin by the APC/GSK3-β complex leads to proteosome mediated degredation. In the presence of Wnt signaling the degredation machinery is blocked leading to the accumulation of cytosolic β-catenin which then interacts with TCF-4 /LEF transcription factors. Nuclear targets are involved in the regulation of the cell cycle and apoptosis. Adapted from Behrens, 2000 and Macleod, 2000.

1.7.5 Mutation of other genes in MSI$^+$ CRCs

TGFBR2 and *BAX* are mutated with consistently high frequency in MSI$^+$ CRCs and various lines of functional and genetic evidence suggests that the p53/BAX and TGFB/SMAD pathways are involved in both MSI$^+$ and MSS tumourigenesis, albeit via different mechanisms of abrogation. Constitutive activation of the β-catenin/TCF pathway is clearly important in MSS CRCs and there is some evidence that its disruption can occur in a mutator pathway of tumourigenesis.

There are a growing number of other genes whose frequency of mutation at short coding repeat sequences in MSI$^+$ CRCs suggests they may also confer a selective advantage (Duval *et al.*, 2001). Such candidates include the PTEN tumour suppresser gene (Guanti *et al.*, 2000), the insulin like growth factor II receptor (IGFRII) (Souza *et al.*, 1996; Ouyang *et al.*, 1997) and the candidate DNA repair gene MBD4 (Bader *et al.*, 1999; Riccio *et al.*, 1999). Interestingly, frameshift mutations within an $(A)_8$ repeat in *MSH3* and a $(C)_8$ tract in *MSH6* have been demonstrated exclusively in around 30-40% of MSI$^+$ CRCs (Malkhosyan *et al.*, 1996; Yamamoto *et al.*, 1999; Cunningham *et al.*, 2001). This has led to a hypothesis in which the mutator phenotype unfolds in a series of consecutive steps. Initial inactivating mutations in MMR genes such as *MLH1* and *MSH2* result in MSI leading to the mutation of other genes containing repetitive tracts that are susceptible to mutation. The somatic inactivation of secondary MMR genes such as *MSH3* and *MSH6* may subsequently contribute to genomic instability by enhancing or broadening this phenotype (Figure 1.4) (Malkhosyan *et al.*, 1996; Yamamoto *et al.*, 1999).

There is considerable variation in the frequency with which candidate genes are mutated and widespread intra-tumoural heterogeneity has been detected in MSI$^+$ CRCs with respect to the spectrum of genes that are mutated (Barnetson *et al.*, 2000). Even non-coding repeated sequences show great variability in their mutation frequency in MSI$^+$ cells, suggesting that the inherent stability of a sequence may significantly contribute to its mutation frequency in the presence of MMR defects (Zhang *et al.*, 2001).

1.8 Research Aims

The central theme of this project was to define the molecular consequences that accumulate exclusively due to defects in MMR and to identify factors that may influence the manifestation of the resultant mutator phenotype. In particular to address whether the frequency of mutations detected in CRCs with MMR defects may be influenced by inherent instability at both non-coding and coding sequences

In view of the evidence in the literature, it was hypothesised that sequences frequently mutated in MMR deficient CRCs, may be inherently prone to mutation. This could explain the heterogeneity in mutation frequency observed at repetitive tracts in both non-coding and coding sequence. If the inherent stability of a given sequence significantly contributes to its mutation frequency in MSI^+ cancers, this would suggest that an increased mutation rate consequent of MMR deficiency, represents a driving force in tumourigenesis. Factors independent of the tumourigenic process itself, which may affect the inherent stability of repeat sequences, would consequently be expected to affect the manifestation of the mutator phenotype. Mutations in the coding sequences of genes, implicated in the progression of MSI^+ tumourigenesis, are thus predicted to occur exclusively as a consequence of MMR defects even when selection pressures are minimised.

It was specifically aimed to investigate the inherent stability of sequences mutated frequently in CRC. Initially at non-coding microsatellite repeats D2S123 and BAT-40 due to their relevance in the assessment of MSI status in colorectal tumours. Differences in inherent stability between different marker loci were predicted to give insight into factors that can affect or modify susceptibility to mutation in the presence of MMR defects. It was also aimed to investigate whether coding sequence mutations frequently detected in CRC arise exclusively when MMR is defective as a result of failure to repair spontaneous mutations. This was with a view to understanding the contribution from inherent instability at such sequences in tumour progression. Specific investigation was carried out at repeat loci in the *TGFBR2* and *BAX* genes due to the high frequency with which they are mutated in MSI^+ CRCs. In addition, inherent stability of regions within the *APC* gene and *CTNNB1* gene in the

presence of MMR defects was also analysed to address whether MMR deficiency contributes to their mutation in MSI^+ CRCs.

MMR defects are associated with a substantial proportion of CRCs as well as cancers of extracolonic origin. Understanding of the molecular basis that underlies these cancers is clinically important in terms of revealing potential therapeutic targets and novel rational for preventative strategies. Furthermore, since cancers with MMR defects display an elevated mutation rate, elucidation of the events that arise exclusively as a result of increased genomic instability may provide new insight into the contribution that an elevated mutation frequency makes to tumourigenesis. This may shed light on the relative balance between selection and inherent instability in the initiation and progression of cancer.

1.9 Experimental Approach

The research presented in this thesis specifically utilises two cell lines, lbl-1260 and lbl-1261. These cell lines are derived from normal tissue and have been previously characterised and demonstrated to be completely deficient in MMR as a consequence of constitutive heterozygous dominant negative MMR mutations (Parsons *et al.*, 1995a). Since lbl-1260 and lbl-1261 are non-cancer derived but have constitutional defects in MMR they provide an excellent system with which to investigate molecular events that arise as a result of such defects and thus address the hypothesis and aims outlined above.

Cancer cell lines are invariably subjected to selection pressures for mutations that provide the cell with a growth advantage, resulting in successive waves of clonal evolution. Secondary defects that may contribute to the mutation rate and additional molecular variability's also occur consequent of tumourigenesis itself. These factors bias the apparent frequency of mutations, making the dissection of events dependent solely on MMR defects difficult. In cells derived from normal tissue these factors are minimised and the presence of MMR defects can then effectively "unmask" inherent stability at given sequences.

In this thesis, lbl-1260 and lbl-1261 are shown to be derived from normal tissue and further evidence is provided that confirms the presence of the MMR defects demonstrated previously (Parsons *et al.*, 1995a). This analysis is presented in Chapter 3. These cell lines are then utilised in the following chapters to address specific questions relating to the hypothesis. While Chapters 4 to 6 focus on the consequences of MMR defects at non-coding microsatellite sequences, Chapters 7 and 8 focus on the effects of such defects at coding sequences within genes implicated in the initiation and development of MSI^+ CRC.

In Chapter 4 the MMR defects in lbl-1260 and lbl-1261, derived from normal tissue, are shown to be associated with low level MSI similar to that observed in MMR defective cancers. This is revealed, by employing a small pool PCR approach to detect mutations in individual alleles. The MSI phenotype is shown to display heterogeneity on a number of different levels and the data indicates that factors independent of tumourigenesis, contribute to inherent stability of the specific repeat sequences analysed. These factors are investigated in Chapters 5 and 6.

Data presented in Chapter 4 reveals striking heterogeneity in the susceptibility of individual alleles to mutation at the microsatellite marker D2S123. In view of this finding, factors that may contribute to allele bias in mutation at D2S123 are investigated in Chapter 5. Individual alleles at this locus have been characterised and their inherent susceptibility to mutation consequent of MMR defects investigated. Allele specific determinants of instability at this locus have been identified and in addition, patient constitutional genotype is specifically demonstrated to affect the manifestation of MSI in MMR deficient tumours.

Observations in Chapter 4 also suggest that the poly(A) marker BAT-40, is particularly prone to mutation in the presence of MMR defects. In Chapter 6 the notion that BAT-40 may be frequently mutated in MMR deficient cells due to its extreme susceptibility to mutation, is further investigated by analysis of this locus in the germline. This is carried out by study of population allele frequencies, mutation frequency in families and mutation frequency in sperm DNA.

Chapter 7 addresses whether differences in the mutation frequencies observed at the repetitive tracts within the *TGFBR2* and *BAX* genes in MSI^+ CRC, may be due to

differential stability in the presence of MMR defects. The repeat regions within these genes, frequently mutated in MSI^+ CRCs, are analysed to examine whether low level mutation can be detected in cells that are not derived from cancer tissue but that have defects in MMR. This analysis is carried out using a combination of cloning, SP-PCR and restriction digest based methods to investigate the mutational status of individual alleles.

Finally an extensive appraisal of the inherent stability of β-catenin/TCF pathway genes in the presence of MMR defects is presented in Chapter 8. This chapter addresses whether the *CTNNB1* and *APC* genes are inherently prone to mutation in the absence of MMR activity. Regions of these genes, mutated frequently in MMR proficient cancers, are analysed in lbl-1261 and lbl-1260 using a cloning based approach to determine whether mutations can occur exclusively as a consequence of MMR defects.

Although detailed discussion of the results is provided in each relevant chapter, Chapter 9 summarises the main themes and conclusions that have emerged during the entire course of this research.

Chapter 2

Materials and Methods

The following chapter documents the methods used during the course of this thesis. Where appropriate, specific detail is included in methodological overviews in the relevant result chapter. Where standard protocols have been used, these have been referred to, and any adjustments made to the cited method have been documented in the text. Standard safety procedures and COSH regulations were adhered to. Stock solutions and media marked with an asterisk (*) were prepared by the solution and media preparation service at the MRC HGU and sterilised by autoclaving. Where the pH of solutions was adjusted this was done by adding conc. HCl or NaOH as appropriate and monitoring pH using a microprocessor pH meter (Hanna Interments).

2.1 Biological Material

2.1.1 Summary of cell lines

The following cell lines were used during the course of this PhD; Epstein-Barr virus (EBV) transformed lymphoblast cell lines lbl-1261 and lbl-1260 were a donation from Prof. Vogelstein (John Hopkins Oncology Centre, Maryland). These cell lines were previously established from two patients with CRC. Both had a family history of the disease as well as phenotypic evidence of Turcots syndrome. Cell line lbl-1261 harbours a dominant negative mutation in the human *PMS2* gene. (Hamilton *et al.*, 1995; Parsons *et al.*, 1995a; Nicolaides *et al.*, 1998). Cell line lbl-1260 has a germline *MLH1* gene mutation that is speculated to be dominant negative (Parsons *et al.*, 1995a). Lbl-1261 and Lbl-1260 are described in more detail in Chapter 3.

Control EBV transformed lymphoblast cell lines (lbl-c5, lbl-c1, lbl-a, lbl-c8), were established previously by Sheila Mcbeath (MRC, Human Genetics Unit, Edinburgh) from healthy anonymous donors. These control cell lines are maintained as laboratory stocks.

All colorectal carcinoma cell lines are available from the American/European Type Culture Collections (ATCC/ECACC).

Table 2.1 Summary of cell lines.

Cell line	Cell type	Mutation	Reference
Ibl-1260	EBV transformed lymphoblast	*MLH1* see 2.1.2	(Parsons *et al.*, 1995a)
Ibl-1261	EBV transformed lymphoblast	*PMS2* see2.1.2	(Parsons *et al.*, 1995a)
Ibl-a	EBV transformed lymphoblast	Control	Lab stock
Ibl-c5	EBV transformed lymphoblast	Control	Lab stock
Ibl-c1	EBV transformed lymphoblast	Control	Lab stock
Ibl-c8	EBV transformed lymphoblast	Control	Lab stock
LoVo	Colorectal carcinoma	*MSH2*	(Liu *et al.*, 1995b)
SW480	Colorectal carcinoma	*APC* and *p53*	(Leibovitz *et al.*, 1976)
HCT116	Colorectal carcinoma	*MLH1*	(Papadopoulos *et al.*, 1994)
HCT116 + chr3	Colorectal carcinoma	Restoration of MMR activity due to Introduction of wild type *MLH1* allele on chromosome 3	(Koi *et al.*, 1994)

2.1.2 Maintenance of EBV transformed lymphoblast cell lines

Media, solutions and additives;

Freezing medium

8% w/v Dimethysulfoxide (DMSO) in new-born calf serum

Tissue culture medium

RPMI (Gibco BRL)

10% w/v Foetal calf serum (FCS)

1% w/v Penicillin and streptomycin

Phosphate Buffered Saline (PBS)*

0.1M $NaH_2PO_4.H_2O$

0.1M $NA_2HPO_4.7H_2O$

pH7.4

Trypsin Versene (TV)*

50% w/v Trypsin

50% w/v Versin

Maintenance of EBV transformed cell lines was carried out by Sheila McBeath and myself. All tissue culture was carried out under sterile conditions in a class 2-containment hood.

In order to grow cells for DNA extraction, protein purification or FACS analysis, cell lines were defrosted by rapid thawing in warm water followed by direct transfer to 5mls of RPMI tissue culture medium and then fed as required. To retain the cell lines as renewable sources at least 3 x 10^6 were split from the main culture, centrifuged at 1200rpm in a Wifug lab centrifuge and the cell pellet re-suspended in 1.5ml freezing medium. The cells were then sequentially frozen at –70°C and –140°C.

2.1.3 Maintenance of adherent cell lines

Adherent cell lines were maintained as for EBV transformed lymphoblast cell lines. Cells were detached from culture flasks using either a cell scraper or by washing with PBS and incubating with 1-3ml TV for 3 minutes. RPMI culture medium was then added accordingly.

2.1.4 Semen samples

Sperm sample MD-c1 was donated anonymously from a healthy control individual. Sperm sample MD-949 carries a germline mutation in the mismatch repair gene, human *MLH1* resulting in a deletion of exon 12 (codons 347-470). Constitutional (blood) samples were also donated from both individuals.

2.2 DNA and RNA Purification Protocols

2.2.1 Purification of DNA from cell lines and blood lymphocytes

Isolation of genomic DNA from cell cultures and blood was carried out using a Nucleon II extraction kit (Scotlab Bioscience, Strathclyde) according to the manufacturer's instructions. This procedure was carried out in a class 2-containment hood

2.2.2 Purification of DNA from sperm

Solutions;

20 x Salt and Sodium citrate buffer (SSC)*

3M NaCl

3.3M Sodium citrate $-2H_20$

pH7.0

10% Sodium dodecyl sulphate (SDS)*

10% w/v SDS crystals

pH7.2

DNA was extracted from sperm as described by Jeffrey's *et a*l (Jeffreys *et al.*, 1990) in a biological class 2 cabinet. Semen was pelleted by spinning at 1200rpm in a Wifug lab centrifuge for 10min. Semen pellets were rinsed 3 times with 20ml 1x SSC followed by 6 washes with 20ml 1x SSC and 1% SDS to lyse any seminal leukocytes and epithelial cells. Centrifugation after each wash was for 5 min at 3500rpm in a Wifug centrifuge. The residual sperm pellet was incubated in 1x SSC and 1M 2-mercaptoethanol (Sigma) at room temperature for 5 minutes and the reduced sperm lysed by addition of SDS to 1%. Sperm DNA was collected after phenol chloroform extraction and ethanol precipitation.

2.2.3 Phenol chloroform extraction and ethanol precipitation

Solutions;

1M Tris(hydroxymethyl)aminomethane (Tris)*

1M (Tris)

pH7.7

0.5M Disodium ethylenediaminetetraacetate-$2H_2O$ (Na_2EDTA)*

0.5M Na_2EDTA

pH8.0

1 x Tris EDTA (TE)*

1M Tris (pH 7.7),

0.5M Na_2EDTA (pH8.0)

3M Sodium acetate (NaOAc)

3M NaOAc-$3H_2O$

pH4

Phenol Chloroform extraction was carried out in a biological class 2 cabinet. An equal volume of phenol chloroform solution (Gibco BRL) was added to the DNA solution. This was inverted and spun for 5 minutes at 2000rpm in a Wifug lab centrifuge. Supernatant containing the DNA was carefully removed with a pipette. An equal volume of chloroform was then added to the supernatant, inverted and spun for 5 min at 2000rpm. The supernatant was again removed.

For ethanol precipitation a 1:10 volume of 3M NaOAc (pH4) was added to the DNA solution along with 2-3 volumes of 100% ethanol. The solution was incubated for 30min at –70°C and then centrifuged for 15min at 2000rpm a Wifug lab centrifuge. The residual DNA pellet was re-suspended in an appropriate volume of 1 x Tris EDTA pH7.7.

2.2.4 Purification of DNA from bacterial plasmids

Plasmid DNA was extracted on small scale using a QIAprep miniprep kit (QIAGEN Ltd, Crawley, UK) according to the manufacturers. For DNA extraction from a large number of plasmid samples (in 96well plates), a BioMek® 2000

workstation robot was used, operated by Dr. Stewart Mckay (MRC, Human genetics Unit, Edinburgh) according to manufacturers instructions.

2.2.5 Estimation of DNA concentration

5µl of genomic and sperm DNA samples were run on an agarose gel at stock concentration and 1:10 dilution (see 2.4.6 for gel electrophoresis) against DNA standards of known concentration (1µg/µl, 750ng/µl, 500ng/µl, 250ng/µl, 100ng/µl). To confirm the concentration indicated by agarose gel electrophoresis, DNA was then measured by optical densomitery on a UV spectrophotometer (Pharmecia). DNA samples were diluted in triplicate and placed in quartz cuvettes. The absorbency at 260nm and 280nm was then measured.

DNA concentration was calculated from the readings as follows;

DNA µg/ml = A^{260} x dilution factor x 50

The A^{260}/ A^{280} ratio of 1.8 taken as optimum purity of DNA. Readings were taken in triplicate for each sample and repeated if inconsistent. Following determination of concentration, stock DNA samples were then diluted to 100-200ng/µl or where starting concentration was less than this, (in the case of sperm DNA), to an appropriate concentration for handling.

2.2.6 Purification of RNA from cell lines

Isolation of RNA from cell cultures was carried out using an RNeasy mini kit (QIAGEN Ltd, Crawley, UK) according to the manufacturer's instructions.

2.3 DNA Samples

The following were obtained directly as DNA samples. Where samples are prefixed with MD- this represents the laboratory identification code used.

2.3.1 CEPH family DNA

CEPH family DNA was obtained from the Foundation Jean Dausset Centre D'Etude Du Polymorphisme Humain by Prof. Alan Wright (MRC, Human Genetics Unit, Edinburgh). DNA from CEPH families was provided in microtiter plates containing lypholized DNA. The families used in this study were randomly selected and were; 66, 1331, 1341, 1346, 1362, 1377, 1423, 13293, and 13294 as indicated by CEPH family identification numbers. Each sample was re-suspended to a final concentration of 100ng/µl in dH_2O and shook gently for 24 hours.

2.3.2 Constitutional DNA from a Scottish cohort

DNA samples of a Scottish cohort were available as laboratory stocks. These originated from the blood of 100 unrelated Scottish individuals. Prof. Malcolm Dunlop (MRC, Human Genetics Unit, Edinburgh) collected these samples from healthy individuals and they were labelled anonymously. They were all diluted and used at 100ng/µl concentration.

2.3.3 Colorectal tumour DNA

Prof. M. Dunlop and Dr. S. Farrington (MRC, Human Genetics Unit, Edinburgh) had previously collected a panel of colorectal tumour DNA samples. These were available as a laboratory resource. Analysis of tumour MSI status had also been determined by Dr. S. Farrington previously, by comparison of matched normal and tumour DNA using a panel of 8 microsatellite markers that included the recommended panel of 5 (Rodriguez-Bigas *et al.*, 1997; Boland *et al.*, 1998). Tumour DNA samples were diluted to a concentration of 100ng/µl for use.

2.3.4 Constitutional DNA from colorectal cancer patients

Prof. M. Dunlop and Dr. S. Farrington collected a panel of constitutional DNA previously, from the blood of patients with CRC. These were available as a laboratory resource. These DNA samples were diluted to approximately 100ng/µl for use.

2.3.5 Constitutional DNA from a Scottish HNPCC family (K-453)

Constitutional DNA was available from laboratory stocks of a Scottish Family, K-435. This family had been previously identified as an HNPCC kindred with affected individuals displaying tumour instability. Dr. S. Farrington has previously confirmed family relationships by genotyping, using a panel of microsatellite markers. DNA samples diluted to 100ng/μl were available from 20 individuals from this family.

2.4 PCR Protocols

Manipulation of oligonucleotides and all PCR techniques were carried out in a biological class 2 cabinet. Pipettes specifically designated for PCR usage were also employed to minimise the possibility of contamination of PCR products.

2.4.1 Oligonuclotides for PCR

Primers were supplied by Genosys as precipitates and re-suspended in dH_2O to a stock concentration of 1μg/μl. Amplification was performed using an omnigene PCR system thermal cycler (Hybaid) under the following standard conditions unless stated otherwise: 94°C for 3min for 1 cycle, 94°C for 1min, 55°C for 1min, 72°C for 1min for 35 cycles, 72°C for 5min for 1 cycle.

Primers were specifically designed unless otherwise referenced (Table 2.2). For certain experiments, primers were fluorescently labelled with 6-carboxyfluorescein (FAM) or 4, 7, 2', 4', 5', 7'-hexachloro-6-carboxyfluorescein (HEX) and in such instances this is indicated in the appropriate results chapter. F and R refer to forward and reverse respectively.

Table 2.2 Oligonucleotides used for PCR

Primer	Sequence 5'-3'	Cycling Conditions	Reference
MLH1-ex15F	CTACCAGATACTCATTTATG	See 2.4.3	Farrington,
MLH1-ex18R	CATTCCTTTTCTTCGTCCC		pers. com.
D2S123F	AAACAGGATGCCTGCCTTTA	standard	(Parsons,
D2S123R	GGGGACTTTCCACCTATGGGAC		1995b)
BAT-40F	AATAACTTCCTACACCACAAC	standard	(Liu *et al.*,
BAT-40R	GTAGAGCAAGACCACCTT		1995a)
TGFBassayF	CACTCTAGGAGAAAGAATGACG	anneal at	(Mironov *et al.*,
TGFBassayR	GAAAGTCTCACCAGGCTTTTTGATT	55-60°C	1999)
TGFBR2ex3F	CCTCGCTTCCAATGAATCTC	standard	(Lu *et al.*,
TGFBR2ex3R	TTGGCACAGATCTCAGGTCC		1996)
TGFBR2ex4F	CCACGTGTGCCAACAACATCAACC	standard	Farrington
TGFBR2ex4R	CAGCCGTCAGGAACTGGAGTA		pers.com.
BAXex3F	ATCCAGGATCGAGCAGGGCG	anneal at	(Rampino *et*
BAXex3R	ACTCGCTCAGCTTCTTGGTG	60°C	*al.*, 1997)
APCex10F	AAACATCATTGCTCTTCAAATAAC	standard	(Farrington
APCex10R	TACCATGATTTAAAAATCCACCAG		and Dunlop,
			1999)
CTNNB1ex3F	TCGTATTTATAGCTGATTTG	anneal at	Farrington
CTNNB1ex3R	TAATACTCTCTTACCAGCTAC	50°C	pers. Com.
APCex15B1	AAGTACAAGGATGCCAATATTATG	anneal at	(Groden *et al.*,
APCex15D2	TGTTTGGGTCTTGCCCATCTTT	50°C	1991)
APCex15G1	AAGAAACAATACAGACTTATTGTG	1.5min	(Groden *et al.*,
APCex15I2	CCGTGGCATATCATCCCCC	extension	1991)
		step	
M13F	GTAAAACGACGGCCAG	standard	
M13R	CAGGAAACAGCTATGAC		

2.4.2 Standard PCR

PCR reactions were performed in either a final volume of 25µl using the Expand High Fidelity PCR system (Boehringer Mannheim, Germany) or a final volume of 50µl using AmpliTaq (Roche). Final reaction concentrations for the Expand High Fidelity system were 1 x PCR buffer II, 0.2mM dNTPs, 100ng oligonucleotide primer, 100ng DNA and 0.87 Units of Expand High Fidelity PCR mix. Final reaction concentrations for the AmpliTaq system were as for the Expand system but using 1x AmpliTaq buffer instead of Expand buffer II, 2.5mM MgCl and 1.25 units of AmpliTaq. Where PCR analysis was directly from bacterial plasmid DNA, bacterial colonies were stabbed into the PCR reactions on the bench by a bunsen flame using a sterile tooth pick, in place of 100ng DNA. The colony was then stabbed into a storage culture in a 96 well culture plate (see 2.6)

2.4.3 Reverse transcription and PCR amplification (RT-PCR)

RT-PCR was performed on purified RNA using a Promega Access RT-PCR system (Promega) according to the manufacturer instructions. Specific amplification conditions for MLH1 RT-PCR (Chapter 3) using MLH1-ex15F and MLH1-ex18R primers were: 1 cycle of 48°C for 45min, 94°C for 2min, followed by 40 cycles of 94°C for 30s, 60°C for 1min, 68°C for 2min followed by 1 cycle of 68°C for 7 min and 4°C infinity.

2.4.4 Dilution protocol for SP-PCR

An assay was devised to genotype individual alleles, using a Small Pool PCR strategy (SP-PCR). A similar approach has been used previously to detect mutations within populations of wild type alleles (Jeffreys *et al.*, 1994; Monckton *et al.*, 1995; Parsons *et al.*, 1995a; Vilkki *et al.*, 2001).

DNA from cell lines was extensively diluted to a final concentration of 15-20pg per PCR reaction to give up to 3 input molecules of each allele per amplification (assuming 6pg DNA per diploid genome). This was carried out by serial dilution of DNA in dH_2O as follows;

From the concentration of DNA determined as described above (2.2.5), dH_2O was added to a final concentration of 40ng/µl. This was followed by 1:10 serial dilutions giving to give aliquot of 4000pg/µl, 400pg/µl, 40pg/µl. This was followed by 1:2 dilution to 20pg/µl. Prior to aliquoting for each dilution, DNA was repeatedly inverted to ensure homogenisation throughout the sample. Limiting dilution SP-PCR reactions (see 2.4.5) were carried out initially using the 40pg/µl and 20pg/µl dilution aliquots in approximately 25 reactions to check the accuracy of the dilution procedure. If SP-PCR products were detected in too many reactions (>40%) the 20pg/µl dilution was further diluted to 15pg/µl and re-tested in preliminary SP-PCR reactions. Due to the dilute nature of the final 15-20pg/µl aliquot some PCR reactions contained no product. Final dilution concentrations resulted in the detection of alleles in ~30% of SP-PCR reactions by mutation analysis suggesting that most products represented single molecules (Parsons *et al.*, 1995a; Vilkki *et al.*, 2001). This was further indicated by the detection of single alleles by mutational analysis of SP-PCR products amplified from individuals who were constitutionally heterozygous at the given locus (See Chapter 4).

2.4.5 Small pool PCR

PCR amplifications were performed in a final volume of 25µl. Final reaction concentrations were as described in 2.4.2 but using 1µl (15pg) of diluted DNA rather than 100ng DNA. Avoidance of contamination was paramount when amplifying dilute DNA. Therefore all reactions were prepared in a class 2-containment hood. All pipettes and plastics used in preparation of the SP-PCR reactions were UV irradiated for 20 minutes in a Template Tamer (Oncor). Buffer solutions and sterile water were opened under sterile conditions and also subject to UV irradiation. All Reactions were prepared in 96 well plates. DNA free controls were prepared in 16 of the wells in each plate (columns 3 and 10) and positive controls containing 100ng-cell line DNA were prepared in 2 wells in every plate. Primers and cycling conditions are given in 2.4.1.

2.4.6 Gel electrophoresis

Solutions

10 x Tris-Acetate EDTA (TAE)*

2M Tris

5.7% w/v Glacial acetic acid

50mM Na_2EDTA (pH8.0)

10 x Tris-Borate EDTA (TBE)*

1.1M Tris

1.1M Boric acid

20mM Na_2EDTA (pH8.0)

Loading Buffer

100mM Na_2EDTA (pH8.0)

0.25% w/v Bromophenol blue

30% w/v Sucrose.

Standard agarose gels were prepared using routine electrophoresis grade agarose (Flowgen) and 1x TAE electrophoresis buffer. Separation of large fragments required low % gels (typically 1.5%) whereas smaller fragments were separated on higher % gels (typically 3%). If DNA fragments were being isolated from the gel for sequencing, a 3% Nusieve GTG agarose (Bioproducts, Rockland, ME), 1% 'Hi Pure' low EEO agarose (Biogene, Kimbolton, UK) gel was prepared using 1x TBE electrophoresis buffer.

5-10µl standard PCR products were loaded onto the gel with 2µl of loading buffer. Size markers used were generally a 1kb ladder (Gibco BRL) or a 100bp ladder (Promega). The DNA was electrophoresed at 30-60V until the dye front was 1cm from the end of the gel although this varied with the size of fragments being investigated. Visualisation of the DNA was by staining with ethidium bromide (Biorad) followed by UV trans-illumination using a Herolab trans-illuminator (Herolab, Wiesloch). Images were captured using a Herolab Camera and Easywin 32 version 2.00 software.

2.5 Mutation Analysis Protocols

2.5.1 Fragment size analysis using an ABI310 genetic analyser

PCR products and SP-PCR products were size analysed using an ABI310 automated genetic analyser with Genescan 350 data collection software version 1.0.4. PCRs were carried out using primers labelled with either FAM or HEX to allow detection by ABI310 analysis and are specified in the appropriate result chapters. Detection of fluorescent labels by the ABI310 is very sensitive, allowing efficient visualisation and sizing of SP-PCR products. PCR products were prepared for analysis by adding 11.5µl deionised formamide (PE Applied Biosystems, Cheshire) and 0.5µl TAMRA 350 size standard (PE Applied Biosystems, Cheshire) to 2µl of PCR product. Samples were denatured by heating for 5min at 95°C and then placed on ice for 2min. Samples were run according to standard protocol using module GS STR POP4 (1ml) with virtual filter set C (PE, Applied Biosystems, Cheshire). Injection time was extended from 5 to 10 seconds for analysis of SP-PCR products and negative controls from SP-PCR plates in order to increase sensitivity. For analysis of SP-PCR products, all 16 negative controls were run from every SP-PCR plate analysed and if a product was observed in any negative sample, the entire plate was discarded. Positive controls from every plate were also analysed to ensure reproducibility of the ABI310 profiles. The size of each allele was taken to be the predominant peak in each peak complex as determined from the Genescan analysis data.

2.5.2 Mutation detection using a Transgenomic Wave™ DNA fragment analysis system

Solutions;

Buffer A*

0.1M Triethylammonium acetate (TEAA) solution pH7.0

Buffer B*

0.1M TEAA

25% Acetonitrile (ACN) pH7.0

The Wave™ (Transgenomic Inc., SanJose USA) was employed to detect mutations in PCR and SP-PCR fragments. Mutations are visualised as a characteristic pattern of peaks corresponding to a mixture of heteroduplexes and homoduplexes formed when wild type and mutant DNA are hybridised. This method has been shown to be highly sensitive at detecting over 97% mutations and is especially appropriate for detecting single base pair changes (Giordano *et al.*, 1999; Taliani *et al.*, 2001).

For detection of SP-PCR products by Wave™ analysis, 3µl of the primary SP-PCR product was subjected to a second round of PCR identical to the first but with the number of cycles reduced from 35 to 25.

10-20µl of PCR or secondary SP-PCR products were prepared for analysis by hybridisation to form heteroduplexes. Samples were heated to 95°C for 3min and then cooled to 25°C in stepwise decreases of 1°C over 45 min using a PCR system thermal cycler (Hybaid). The Wave™ instrument was set up for each run according to manufacturer instructions with default settings. Adjustments made were to maximum noise and drift settings in the sample table, setting them at 500. An injection volume of 10µl was used in all runs. Optimal gradients and temperatures were determined for each different fragment being analysed. A universal gradient was run and the parameters adjusted accordingly to give a total run time of 7.3min with fragment elution at approximately 5.3min. A series of melting curves was then performed at 1-2°C increments between 50°C-65°C to determine the optimal temperature for mutation detection. These standard procedures were carried out according to manufacturer instructions. Optimal gradients and temperatures were confirmed by the detection of heteroduplex peaks in samples containing known mutations. The separation gradients and temperatures used for each fragment analysed in this study are given in 2.5.3.

2.5.3 Separation gradients and temperatures for Wave™ mutation detection

A= Buffer A

B= Buffer B

Slope of 2% increase in buffer B per minute

BAX exon 3

For analysis of a 96bp fragment amplified using primers BAXex3F and BAXex3R (2.4.1). Melting temperature; 60°C

Table 2.3 Wave™ Melting gradient for fragment *BAX* exon3.

Time/min	%A	%B
0	63	37
0.1	58	42
4.6	48	52
4.7	0	100
5.2	0	100
5.3	63	37
7.3	63	37

APC exon 10

For analysis of a 227bp fragment amplified using primers APCex10F and APCex10R (2.4.1). Melting temperature; 55°C

Table 2.4 Wave™ Melting gradient for fragment *APC* exon 10.

Time/min	%A	%B
0	53	47
0.1	48	52
4.6	39	61
4.7	0	100
5.2	0	100
5.3	53	47
7.3	53	47

2.5.4 Determination of allele origin

In instances where origin of new alleles was determined (highlighted specifically in the appropriate result chapter) this was done in accordance with the literature e.g. (Weber and Wong, 1993). The origin of "new/mutant" alleles is such that if there were two possibilities, the shortest mutational step was considered to be the actual one. For example if one progenitor allele differed by one repeat and the other by two repeats, when compared to the mutant, a one step mutation was inferred. If two progenitor alleles exhibited the same difference when compared to the new/mutant one, the origin was declared ambiguous.

2.5.5 *TGFBR2* restriction digest assay

A restriction digest based assay was adapted from Mironov *et al,* (Mironov *et al.*, 1999) specifically for the detection of 1bp deletions in the poly$(A)_{10}$ tract of the *TGFBR2* gene (Figure 7.1). The reverse primer introduces a *Hinf1* site in the presence of a 1bp deletion in the poly$(A)_{10}$ tract, whereas in the presence of wild type sequence no *Hinf1* restriction site is introduced during the PCR step (Mironov, *et al.*, 1999). The forward primer is fluorescently labelled with FAM allowing detection of digestion products using an ABI310 genetic analyser. Wild type undigested fragments are visualised at 141bp and mutant *Hinf1* digest fragments at 118bp.

PCR of TGFBR2 assay fragments from genomic DNA was performed as described above using TGFBassayF and TGFBassayR primers (2.4.1) and the Expand High Fidelity PCR system (2.4.2). 5µl of PCR product was digested at 37°C overnight in a total volume of 15µl with 5 units *Hinf1* restriction enzyme and 1 x buffer H (Boehringer Mannheim). Products were analysed on an ABI 310 Automated Genetic Analyser, using Genescan software as described above (2.5.1).

2.6 Cloning and Bacterial Culture

Media;

Luria Broth (L-broth)*

0.1% w/v Tryptone (Difco)

0.05% w/v Yeast extract (Difco),

171mM NaCl

Luria Agar (L-agar)*

0.1% w/v Tryptone (Difco)

0.05% w/v Yeast extract (Difco),

171mM NaCl

0.15% w/v Agar (Oxoid Ltd)

Terrific Broth (Ter-Broth)*

0.12% Tryptone (Difco)

0.24% w/v Yeast extract (Difco)

0.04% w/v Glycerol

1 x TB phosphate

10 x TB phosphate*

0.17M KH_2PO_4

0.72M K_2HPO_4

Additives

Ampicillin stock solution

20mg/ml ampicillin (Sigma)

Isopropyl β-D-thiogalactopyranoside (IPTG) stock solution

100mM IPTG (Sigma)

5-Bromo-4-Chloro-3-Indolyl-β-D-galactosidase (X-gal)

40mg/ml X-gal (Sigma)

2.6.1 TA cloning and transformation

Individual alleles were cloned using a TOPO TA Cloning kit version K2 (Invitrogen BV, Groningen, The Netherlands). Genomic DNA was PCR amplified as described above (2.4.2) using the Expand High Fidelity PCR system (Boheringer

Mannheim). The final extension step in the PCR reaction was extended to 10 minutes to ensure addition of the single deoxyadenosine (A) to the 3' ends of PCR products necessary for efficient ligation into the cloning vector. PCR products were cloned into pCR 2.1-TOPO vectors (Invitrogen BV) according to the manufacturers instructions. Clones were then chemically transformed into either TOP10 or TOP10F' *E.coli* strains (Invitrogen BV) according to manufacturer's instructions. 10-50µl of each transformation was spread onto selective L-agar plates containing amp (50µg/µl) and X-gal (40µg/ml) and/or IPTG (0.2mM) as appropriate. Plates were incubated upturned overnight at 37°C

2.6.2 Colony selection and storage

White and light blue bacterial colonies, indicative of positive transformants, were picked using sterile cocktail sticks or pipette tips. Presence of a correct sized insert was determined by direct PCR amplification using M13F and M13R primers as described above (2.4.1 and 2.4.2) and extending the PCR hot start to 94°C for 6min to burst bacterial walls prior to amplification of plasmid DNA.

Bacteria were stabbed into PCR reaction mixtures after first inoculating a stock culture. For analysis of small numbers of clones, 15ml cultures of L-broth containing 50µg/ml ampicillin were inoculated. These were incubated at 37°C and shook at 225rpm for approximately 16 hours. Stocks for long-term storage were made by mixing 0.85ml of this culture with 0.15ml of sterile glycerol (Sigma). These were transferred to a cryovial and stored at -70°C. DNA was extracted from the remaining culture as described above for small-scale bacterial DNA extraction (2.2.4).

For analysis of large numbers of clones, individual colonies were picked into 96 well plates containing 175µl/well L-broth and 50µg/ml ampicillin. These were incubated at 37°C, and shook at 225rpm for approximately 10 hours. These 96 well stock plates were then stored at –70C after addition of 25µl sterile glycerol to each well. PCR analysis as described above, indicated clones containing an insert and these were cultured for DNA purification. 96 well culture plates containing 1.2mls/well Ter-broth with 50µg/ml ampicillin were inoculated from the original 96 well stock culture. These were incubated at 37°C for approximately 18 hours,

shaking at 300rpm. DNA was then purified from plasmids on large scale using a Biomeck robot as described above (2.2.4).

2.7 Sequence Analysis

2.7.1 Purification of PCR products

PCR products were cleaned for sequencing analysis using a QIAquick PCR purification kit with a microcentrifuge (QIAGEN Ltd., Crawley, UK) according to the manufacturer's instructions. PCR products cut out of agarose gels were cleaned using a QIAquick gel extraction kit (QIAGEN Ltd., Crawley, UK) again, according to manufacturers instructions. In both cases DNA was eluted in 30µl of elution buffer to increase DNA concentration.

2.7.2 DNA sequencing

All sequencing reactions were performed using ABI PRISM Ready Big Dye Terminator cycle Sequencing Kit with AmliTaq DNA polymerase FS (Taq-FS; Perkin Elmer /Applied Biosystems, Cheshire, UK) and Applied Biosystems DNA sequencer models 373A, 377 or 3700 for 96 well format.

Sequencing of purified PCR products was carried out in 10µl reactions using 100ng purified DNA, 40ng primer and 4µl big dye terminator (PE/Applied Biosystems). Amplification was performed on a PCR system thermal cycler (Hybaid) at 96°C for 30 sec, 50°C for 15sec, 60°C for 4 min for 25 cycles.

Sequencing of plasmid DNA was carried out in 22µl reactions using approximately 100ng plasmid DNA, 100ng primer, 4µl Big Dye (Taq-FS; Perkin Elmer /Applied Biosystems, Cheshire, UK) and 0.36mM MgCl. Amplification was performed on a PCR system thermal cycler (Hybaid) at 96°C for 5min for 1 cycle, 96°C for 30sec, 55°C for 15sec, 60°C for 4min for 25 cycles. All sequencing reactions were stored at 4°C prior to precipitation

Sequencing reactions were carried out in either individual epindorf tubes or in 96 well plates.

2.7.3 Precipitation of DNA from sequencing reactions

After amplification, standard precipitation of sequenced DNA was carried out by adding 55µl 95% ethanol and 2µl 3M NaOAC (pH 4) to the sequencing reaction mix. The mixture was incubated on ice for 10min and then spun at 15000rpm in an IEC Micromax centrifuge for 30min. The DNA pellet was re-suspended in 70% ethanol and spun at 15000rpm for a further 10min. Ethanol was removed and the DNA pellet dried at room temperature ensuring complete evaporation of the alcohol. The cell pellet was stored at –20C.

For precipitation of DNA in 96 well format the above procedure was carried out except samples in 96 well plates were incubated at room temperature for 30 mins. Plates were spun in a Sorvell RT6000 centrifuge at 2000rpm for 30 minutes and supernatant was removed by inverting the plate and then, pulse spinning the upturned plates on paper towels at 800rpm. 70% ethanol was added to the cell pellets down the side of each well and rapidly inverting the plate immediately. The pellets were dried by a further pulse spin at 800rpm in a Sorvell RT6000 centrifuge. Cell pellets were again stored at -20°C.

2.7.4 Gel electrophoresis of sequenced DNA

Precipitated cell pellets were re-suspended in 2-4µl loading dye (PE applied Biosystems) and heated at 90°C for 2min by Agnes Gallagher (MRC, Human Genetics Unit, Edinburgh). Samples were then incubated on ice until loading. Samples were analysed using Applied Biosystems DNA sequencer models 373A, 377 or in the case of analyses of samples in 96 well formats an ABI 3700 all according to manufacturer's instructions.

2.7.5 Analysis of sequence data

Raw sequence data was analysed using Sequencing Analysis Version 3.0 (PE Applied Biosystems).

For alignment of multiple sequences of the same fragment, sequence data was imported into the Sequencher program Version 3.0.1 (Gene Codes Corp., Michigan). Up to 100 fragments at a time were aligned with the default assembly parameters

adjusted to allow 75-85% minimum match. The relevant reference sequence was also imported from GenBank and incorporated in alignments to allow comparison of the sequenced fragments to the published sequence.

Alignments were trimmed of vector sequence and poor quality end sequences prior to further analysis. Individual base pairs highlighted by the Sequencher program as showing mutation, deletions or ambiguous sequence were then manually curated.

2.8 Protein Biology

Solutions;

Lysis Buffer

50mM Sodium chloride (NaCl)

10mM N-2 hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES) (pH8.0)

500mM Sucrose

1mM EDTA

0.5mM Spermidine

0.15mM Spermine

0.2% w/v Triton X-100.

Hypertonic Buffer

350mM NaCl

10mM HEPES (pH8.0)

25% w/v Glycerol

0.1mM EDTA

0.5mM Spermidine

0.15mM Spermine.

6x Sample Buffer

20% w/v Glycerol

2% w/v SDS

0.25% w/v Bromophenol blue

1x Stacking buffer

5% w/v β-mercaptoethanol

4x Resolving Buffer

1.5M Tris

0.4% w/v SDS

pH 8.8

4x Stacking Buffer

500mM Tris

0.4% w/v SDS

pH6.8

10x Running Buffer

250mM Tris

2M Glycine

1% w/v SDS.

Semi-Dry Transfer Buffer

47mM Tris

40mM Glycine

0.037% w/v SDS

100mM Methanol.

7% Resolving Gel

1x Resolving buffer

7% w/v Acrylamide

0.15% w/v Ammonium persulphate (APS)

0.01% w/v N, N, N', N' tetramethyl-1-2-diaminomethane (TEMED)

5% Stacking Gel

1x stacking buffer

5% w/v acrylamide

0.15% w/v APS

0.01% w/v TEMED

2.8.1 Preparation of nuclear extracts

Nuclear levels of MLH1 and PMS2 protein were detected by western blotting. Approximately 250 x 10^{-4} confluent cells were used to prepare each extract. Cells were spun at 1200rpm in a Wifug lab centrifuge for 15min. Cell pellets were washed twice in PBS followed by a spin at 1200rpm for 5min after each wash. Nuclear

extracts were prepared by lysing cells in approximately 3x cell volume of lysis buffer (~100µl), Complete™ protease inhibitor cocktail at 1:1250 dilution (Roche Diagnostics, Mannheim, Germany), 1mM pepstatin A (Sigma) and 100mM Pefabloc (Roche Diagnostics). Lysis reaction was blocked on ice, shaking occasionally, for 15min and the cytosolic extract removed by centrifugation at 13000rpm in an IEC micromax centrifuge for 1min. Nuclei were then lysed by re-suspension in approximately 1.5x cell volume (50µl) hypertonic buffer, protease inhibitors and pefabloc as above and incubated on ice for 30min. Debris was cleared by centrifugation at 13000rpm for 5min. Total protein concentration of nuclear extracts was determined by Bradford assays (Biorad, Hercules, Calif) and the nuclear extracts adjusted to the same concentration.

2.8.2 Western analysis

10µg-20µg of nuclear extract was added to a 1:6 dilution of sample buffer boiled for 5min and placed on ice. Samples were resolved by denaturing SDS-PAGE on a 7% polyacrylamide gel in 1 x running buffer. Pre-stained molecular weight markers (Biorad) were run in parallel. Protein was transferred to a nitro-cellulose membrane using a Biorad mini Trans–blot semi-dry Transfer cell for 30 minutes at 10 volts. Prior to transfer, gels were pre-soaked in semi-dry transfer buffer and the nito-cellulose membrane prepared by coating first in 100% methanol and then also in transfer buffer. Successful transfer of proteins was established by covering the nitro-cellulose membranes with Ponceau solution (Sigma). Blots were blocked in 5% dried milk (Marvel) and 0.3% Tween (Sigma) in PBS, overnight at 4°C. Blots were then incubated for 1 hour with monoclonal antibodies PMS2 clone 37 at 1:500 dilution (Transduction laboratories, KY), MLH1 clone G168-15 at 1:250 dilution (Pharmingen Int.) or α-tubulin clone DM 1A (Sigma) at 1:1000 dilution. Blots were washed in 5% dried milk (Marvel) and 0.3% Tween (Sigma) in PBS for 20 min then incubated for 1 hour with horeseradish preoxidase conjugated sheep antimouse IgG secondary antibody. Blots were again washed as previously with a final wash containing just 0.3% Tween in PBS. Protein bands were detected using ECL western blotting protocol (Amersham Little Chalfont, UK) according to manufacturer's instructions.

2.9 Flow Cytometery

2.9.1 Flow cytometry protocol

Assistance with flow cytometry was given by Dr. Martin Waterfall (John Hughe Benett Laboratories, Dept. Oncology, University of Edinburgh).

Cells were harvested, washed twice in PBS (supplemented with 0.1% BSA and 0.1% sodium azide), counted and finally re-suspended at 2 x 10^6 cells/ml for use. Aliquots of 2 x 10^5 cells were incubated for 40mins at 4°C with optimum concentrations of appropriate monoclonal antibody (determined previously by Dr. M. Waterfall by titration), washed twice to remove unbound antibody and re-suspended in PBS for acquisition. Where a second step-detecting antibody was used, incubation and washing steps were repeated as above. Data for 10,000 cells was acquired and analysed using a FACS Caliber and CellQuest software (Becton Dickinson).

The following markers were used in the FACS analyses; PE conjugated anti-human CD19, anti-human CD1a, anti-human CD16, anti-human CD20, anti-human CD80 and anti-human CD83 (Caltag). FITC conjugated anti-human CD3, anti-human CD14, anti-human CD38, anti human CD40, HLA-DR and anti-human CD86 (Caltag). Tricolour conjugated anti-human CD45RO (Caltag). Unconjugated anti-cytokeratin antibody (clone BER-EP4, Dako), detected using a second step FITC conjugated anti-mouse IgG clone (BIOSOURCE, Camarillo, CA, USA).
Excitation of the conjugates is at 488nm. Different emission wavelengths were detected as follows: FL-1, detects emission for FITC at 525nm, FL-2, detects emission for PE at 575nm and FL-3, detects emission for tricolor at 667nm.

2.10 Calculations and Statistics

2.10.1 Statistical analysis

Statistical advice was provided by Dr. Peter Teague (MRC, Human Genetics Unit, Edinburgh). This was primarily in the form of confirming that the most appropriate statistical test was being employed to address any given null hypothesis. Chi squared analyses were carried out to test whether the frequency of a given event was

observed more often than would be expected by chance using the minitab statistical package (Minitab V.13). Where expected counts were less than five, a Fishers Exact test was more appropriate and this was carried out on the following web site http://home.clara.net/sisa/fisher.htm taking a two sided p value for $p(O>E \mid O<E)$ where O= observed and E= expected. To test for significant differences between the distributions of any two non-parametric data sets, a Mann-Whitney U test was performed using the minitab statistical package (Minitab V.13)

Specific statistical calculations and methodology of data analysis is described in more detail in the relevant results chapters.

Chapter 3

Characterisation of Cell Lines Derived from Normal Tissue with Defects in MMR

3.1 Introduction

HNPCC is an autosomal dominant disorder and thus non-cancerous cells of affected patients contain one mutant and one wild type allele (Leach *et al.*, 1993). These phenotypically normal cells are assumed to be MMR proficient since, studies have shown they are microsatellite stable and display no evidence of biochemically measurable MMR deficiency (Parsons *et al.*, 1993; Modrich, 1994). However rare individuals have been identified with congenital defects in MMR due to bi-allelic or dominant negative germline mutations (Parsons *et al.*, 1995a; Ricciardone *et al.*, 1999; Wang *et al.*, 1999; Vilkki *et al.*, 2001)

Lbl-1260 and lbl-1261 are EBV transformed cell lines derived from the B cell lymphocytes of two such patients. These cell lines have been previously isolated and shown to be completely defective for MMR despite being derived from normal tissue (Parsons *et al.*, 1995a). The laboratory of Prof. Bert Vogelstein kindly donated lbl-1260 and lbl-1261, EBV transformed lymphoblastoid cell lines. These cell lines provide an ideal system with which to investigate the consequences of MMR defects at both non-coding and coding sequences in vivo, while minimising the other effects from tumourigenesis that can bias the apparent frequency of mutations.

Cell line lbl-1260 has a germline *MLH1* mutation such that an amino acid is lost through a 3-nucleotide deletion at codon 618 (AAG AAG AAG GCT→AAG AAG GCT, lys deletion). Parsons *et al* refers to lbl-1260 as patient 4 (Parsons *et al.*, 1995a).

Cell line lbl-1261 harbours a germline truncating point mutation at codon 134 in the *PMS2* gene (CGA→TGA, arg 134→stop). Lbl-1261 is patient 6 as referred to by Parsons *et al* (Parsons *et al.*, 1995a; Hamilton *et al.*, 1995; Nicolaides *et al.*, 1998).

Both cell lines have been demonstrated to lack measurable MMR activity by analysis of nuclear extracts (Parsons *et al.*, 1995a). In addition the truncating mutation in lbl-1261 has been shown to exert a dominant negative effect (Nicolaides *et al.*, 1998). In an earlier report the *MLH1* mutation in lbl-1260 was also speculated, to be dominant negative (Parsons *et al.*, 1995a) but this has not been confirmed.

Both patients from whom these cell lines were derived had a family history of CRC. Besides developing CRC they also displayed extra colonic lesions (Hamilton *et al.*, 1995). The male patient from whom cell line lbl-1260 was derived developed CRC by the age of 30 and a gliobastoma at the age of 33 and thus displayed phenotypic characteristics of TS (Hamilton *et al.*, 1995). The female patient from whom cell line lbl-1261 was derived, developed CRC that was detected when she was 11. Phenotypic evidence of adenomatous polyposis was detected at the age of 14 and cutaneous café au lait spots were also evident (Hamilton *et al.*, 1995). Despite phenotypic evidence of adenomatous polyposis, no germline mutations of the *APC* gene were detected in lbl-1261 (Hamilton *et al.*, 1995).

The cell lines lbl-1260 and lbl-1261, which have constitutional defects in MMR and are derived from phenotypically normal tissue, represent an excellent tool with which to analyse the inherent stability of particular sequences. They provide many advantages over investigation using cancer cell lines, which are subject to selection pressures and clonal evolution (Nowell, 1976). In addition, cancer cell lines by their nature accumulate multiple molecular variability's making it difficult to dissect the contribution of events resulting directly and exclusively from MMR defects.

The experiments in this chapter were carried out to further characterise cell lines lbl-1261 and lbl-1260. The lineage of origin of these cell lines was determined in order to demonstrate they are indeed derived from B-lymphocytes. Furthermore, expression of MLH1 and PMS2 proteins was evaluated to gain further insight into the nature of the constitutive MMR defects in these cell lines, particularly in cell line lbl-1260 where the mechanism by which the constitutive *MLH1* mutation confers complete loss of MMR activity, has not been definitively established.

The data presented in this chapter thereby allows lbl-1260 and lbl-1261 to be utilised with confidence in subsequent investigations that address specific questions relating to the consequences of MMR defects.

3.2 Methodological Overview

3.2.1 Maintenance of cell lines

EBV transformed lymphoblastoid cell lines, lbl-1260, lbl-1261 and lbl-a (2.1.1) were maintained as described in 2.1.2. The adherent cancer cell line HCT116 was maintained as described in 2.1.3

3.2.2 Flow cytometry

Dr. Martin Waterfall provided assistance with FACs analysis.

Cell lines lbl-1261, lbl-1260 and control lymphoblastoid cell line lbl-a, were evaluated by FACS analysis as described in 2.9. Cancer cell line HCT116, provided a positive control for epithelial cell marker BER-EP4. The cell surface markers used are described in 2.9.1 and listed in Table 3.1.

Table 3.1 Cell surface markers used in FACS analysis. These markers were used to determine the origin of cell lineage of MMR deficient cell lines, Ibl-1261 and Ibl-1260.

Marker	Cellular Expression
CD1a	Thymocytes, langerhans cells, interdigitating cells
CD3	Thymocytes, T cells
CD14	Peripheral blood monocytes, macrophages, granulocytes
CD16	NK cells and neutrophils, macrophages
CD19	B cells- all stages of differentiation except plasma cells
CD20	B cells- pre, resting and activated except terminally differentiated plasma cells
CD38	Early B and T cells, activated T cells, germinal centre B cells, plasma cells
CD40	B cells, monocytes, dendritic cells
CD45RO	Thymocytes, activated T cells, monocytes, granulocytes. Weakly expressed on activated B-cells and in lymphoblastoid cell lines
CD80	B cell subset, macrophages, dendritic cells, activated T cells
CD83	Activated B cells, putative dendritic cell marker, some germinal centre B cells.
CD86	Activated B cells, monocytes, dendritic cells
HLA-DR	B cells, monocytes, macrophages, activated T cells
BER-EP4	Anti-cytokeretin, epithelial cell marker

3.2.3 Western analysis

Western analysis was carried out as described in 2.8 using monoclonal antibodies PMS2 (Transduction laboratories, KY) or MLH1 (Pharmingen Int.). Monoclonal antibody α-tubulin (Sigma) was also used as a loading control.

Cell lines lbl-1261 and lbl-1260, which harbour constitutive mutations in *PMS2* and *MLH1* respectively, were evaluated for the nuclear expression of both proteins. In addition nuclear extracts were analysed from positive control, MMR proficient, lymphoblastoid cell lines, lbl-a and lbl-c8. In addition, nuclear extracts from a HeLa cell line (Transduction laboratories) provided an additional positive control for PMS2 expression. Cancer cell line HCT116 provided a negative control for loss of MLH1 expression. A negative control for the PMS2 protein was not available. Each analysis was carried out in triplicate. The concentration of nuclear protein loaded onto each gel was evaluated by Bradford Analysis as described in 2.8.1 or by probing the gel with α-tubulin antibody as described in 2.8.2.

3.2.4 RT PCR and cloning

RNA was extracted from cell line lbl-1260 as described in 2.2.6. RT-PCR carried out as described in 2.4.3 using MLH1 primers 15F and 18R (2.4.1) to amplify a 335bp region around the known *MLH1* mutation in lbl-1260 at codon 618. The products were then cloned into TA cloning vectors (Invitrogen) as described (2.6.1 and 2.6.2). Sequencing and analysis of individual alleles was carried out using M13F and M13R primers (2.4.1) as detailed in 2.7

3.3 Results

3.3.1 Cell lines lbl-1260 and lbl-1261 are derived from a B-Lymphocyte cell lineage

Cell surface marker expression of lbl-1260, lbl-1261 and lbl-a was analysed by flow cytometery using the panel of markers described in Table 3.1. Cancer cell line HCT116 was analysed specifically as a positive control for epithelial cell surface marker BER-EP4.

To analyse the FACS data, gates were set on unstained cell populations and applied to subsequent analyses. This minimises distortion of the data from cell auto-fluorescence. Gated cells represent approximately 70% of the total, in all cases. Representative FACS profiles are shown in Figure 3.1 and the data for all analyses is given in Table 3.2.

CD19 and CD20 are classic markers for determining B cell lineage and were detected on >99% gated cells in each of the lymphoblastoid cell lines (Figure 3.1B and Table 3.2). Cancer cell line, HCT116 was also tested for expression of CD19 and as expected CD19 expression was not detected (Table 3.2). All three lymphoblastoid cell lines were positive for five other B cell associated markers (CD38, CD40, CD80, CD86 and HLA-DR) (Figure 3.1B, C, G, H, K-L and Table 3.2). Furthermore, dual staining with CD-19 and CD-20 (Figure 3.1K+L) demonstrated that approximately 100% of the cell population express both markers to the level of detection.

The cell lines were also analysed using markers specific to Thymocytes and T cells (CD1 and CD3), to blood monocytes and macrophages (CD14, CD16) and to NK cells and neutrophils (CD16). These markers provided a panel of negative controls (Table 3.1). No expression of these markers was detected in the lymphoblastoid cell lines (Figure 3.1D, E, I and J and Table 3.2). CD45RO and CD83 were weakly expressed on all three lymphoblastoid cell lines (Table 3.2). Expression of these markers is generally characteristic of activated T cells and dendritic cells. However, germinal centre B cells and lymphoblastoid cell lines have been reported to express low levels (Waterfall, pers. comm.). Expression in a subset

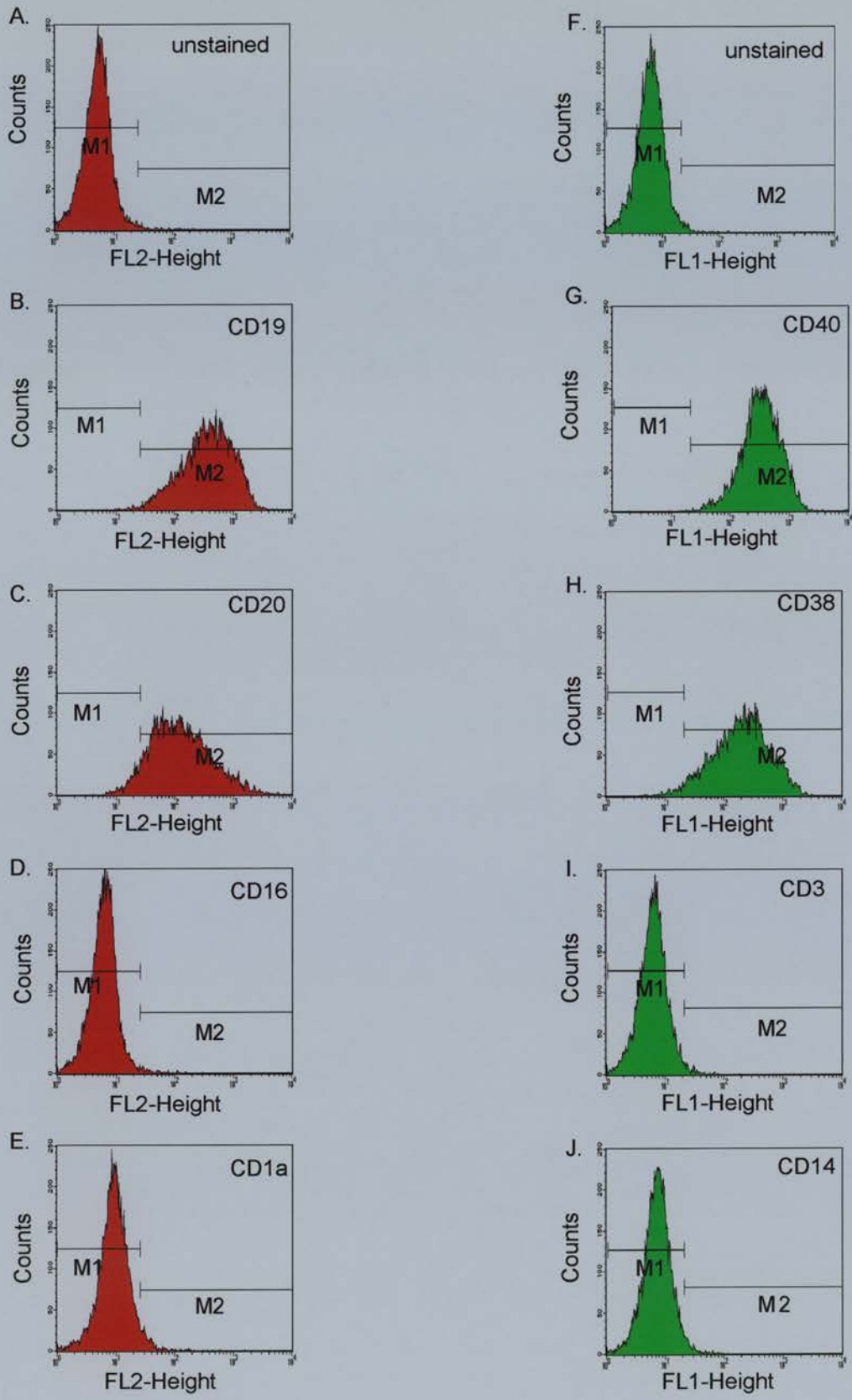
of the lymphoblastoid cell populations as seen here, is entirely consistent with them being of B-lymphocytic origin.

Of further importance, the three lymphoblastoid cell lines were negative for epithelial (anti-cytokeratin) cell specific marker BER-EP4 (Figure 3.1M-N and Table 3.2). Epithelial cancer cell lines, such as those derived from primary CRCs, would be expected to be positive for expression of this marker. Indeed, expression of BER-EP4 was detected on approximately 100% of HCT116 cells (Figure 3.1O-P). The fact that expression of an epithelial cell marker was not detected in the three-lymphoblastoid cell lines indicates they are not of primary tumour origin.

This comprehensive analysis of cell surface marker expression, reveals that cell lines lbl-1261, lbl-1260 and control cell line lbl-a are derived from B-cell lymphocytes, and display no evidence of contamination with epithelial tumour cells.

Table 3.2 Summary of FACs analysis on 2 MMR deficient lymphoblastoid cell lines (lbl-1261 and lbl-1260), 1 MMR proficient lymphoblastoid cell line (lbl-a) and a MMR deficient cancer cell line (HCT116). The percentage of gated cells that were positive for the expression of each marker is given and whether marker expression was positive (+) negative (-) or weakly expressing (w) in each cell line is indicated. Peak channel values refer to the intensity of the positively expressing cells. B-cell specific markers are highlighted in red and markers are divided according to the wavelengths of emission at which they are detected. Unstained cells (neg) indicate the level of auto-fluorescence detected at each wavelength.

Marker		**lbl-1261**			**lbl-1260**			**lbl-a**			**HCT-116**	
	+/-	Positive gated (%)	Peak channel	+/-	Positive gated (%)	Peak channel	+/-	Positive gated (%)	Peak channel	+/-	Positive gated (%)	Peak channel
FITC-conjugate												
Neg.	-	1.0	28	-	1.2	27	-	3.0	27			
CD40	+	99.8	378	+	99.8	273	+	99.7	453			
CD38	+	96.6	228	+	96.7	184	+	99.6	294			
HLADR	+	91.1	437	+	95.3	437	+	91.6	1000			
CD86	+	89.8	86	+	85.8	128	+	85.5	191			
CD3	-	1.1	31	-	2.8	25	-	3.1	26			
CD14	-	1.6	25	-	2.6	25	-	3.5	25			
PE-conjugate												
Neg.	-	1.0	30	-	1.1	25	-	2.7	32	-	0.5	31
CD20	+	90.7	62	+	97.8	1286	+	94.1	83			
CD19	+	99.4	504	+	99.7	697	+	99.8	294	-	1.0	29
CD80	+	85.6	80	+	85.2	56	+	94.3	143			
CD83	w	54.7	30	w	54.2	32	w	59.8	30			
CD1a	-	4.9	25	-	5.4	25	-	7.5	25			
CD16	-	1.3	25	-	2.4	25	-	4.05	25			
Tricolor-conjugate												
Neg.	-	1.6	17	-	1.5	17	-	3.3	18			
CD45-RO	w	30.4	18	w	77.6	39	+	97.45	77			
FITC-secondary stain												
neg	-	0.2	17	-	1.0	17	-	0.4	19	-	0.3	17
BER-EP4	-	1.0	17	-	0.3	19	-	2.1	17	+	98.0	177


A.
unstained
Counts
M1
M2
FL2-Height
F.
unstained
Counts
M1
M2
FL1-Height
B.
CD19
Counts
M1
M2
FL2-Height
G.
CD40
Counts
M1
M2
FL1-Height
C.
CD20
Counts
M1
M2
FL2-Height
H.
CD38
Counts
M1
M2
FL1-Height
D.
CD16
Counts
M1
M2
FL2-Height
I.
CD3
Counts
M1
M2
FL1-Height
E.
CD1a
Counts
M1
M2
FL2-Height
J.
CD14
Counts
M1
M2
FL1-Height

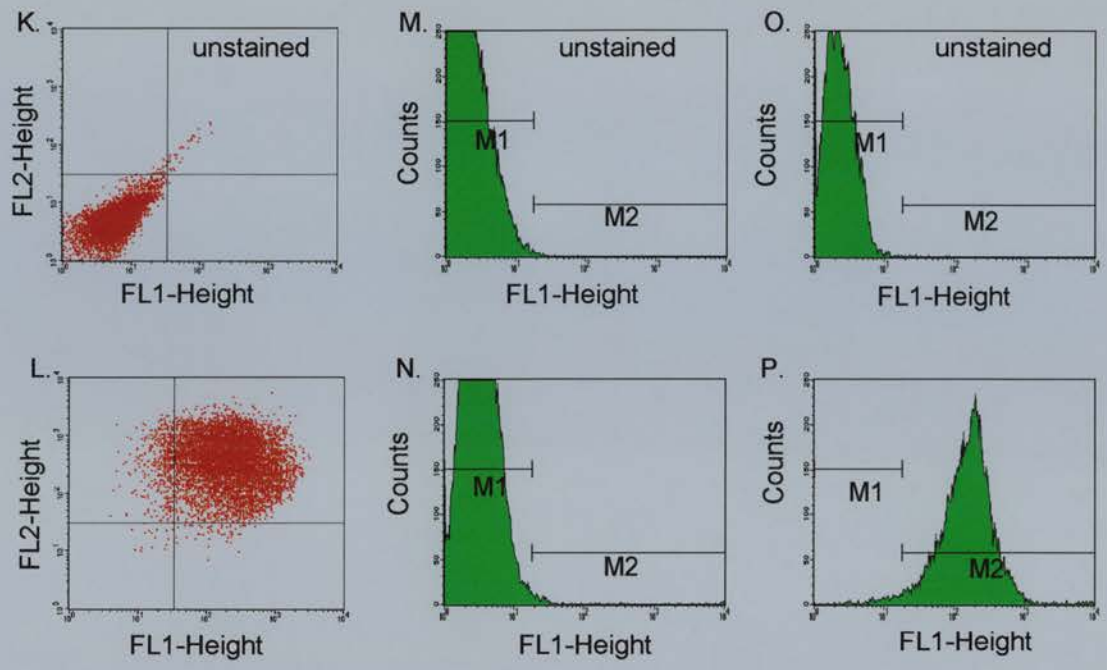


Figure 3.1 Representative histograms and scattergrams from FACS analysis cell surface marker expression. All analyses with the exception of O and P show expression of cell suface markers in MMR deficient cell line Ibl-1261. A and F show unstained gated cells detected in the wavelength of emmission for PE (FL-2 height) or FITC (FL-1 height). Subsequent histograms show detection of a panel of B-cell specific (B-C, G-H), and B-cell negative (D-E, I-J) markers. K and L show Ibl-1261 cells unstained (K) and dual stained (L) with CD38 (FL-1) and CD19 (FL-2). The cells in the upper right gate (L) represent those in which expression of both markers is detected. Analysis of cells for epithelial cell marker expression is shown in M-P. Comparison to unstained Ibl-1261 (M) and HCT116 (O) cells respectively indicates that Ibl-1261 cells are negative for BER-EP4 expression (N) whereas cancer cells, HCT116 clearly express this marker (P). M1 is the marker set to identify the % of unstained cells for the appropriate emmision and M2 is set to identify the % of positively stained cells.

3.3.2 Abnormal MMR gene expression in lbl-1260 and lbl-1261

The constitutive mutations harboured by the patients from which cell lines lbl-1260 and lbl-1261 were derived, have been determined previously (Hamilton *et al.*, 1995). The *PMS2* mutation in cell line lbl-1261 results in truncation of the wild type protein from 862 to 134 amino acids (Nicolaides *et al.*, 1998). Cell line lbl-1261 was analysed for expression of PMS2 using an antibody that specifically detects the wild type protein. As expected, expression of full length PMS2 is reduced by half, in lbl-1261 compared with positive control cell lines (Figure 3.2A) supporting previous analyses of the *PMS2* defect in lbl-1261 (Hamilton *et al.*, 1995; Nicolaides *et al.*, 1998). Lbl-1260 also showed approximately twice the level of PMS2 expression compared to lbl-1261 (Figure 3.2A).

The underlying cause of the MMR defect in lbl-1260 has not been concluded functionally and thus western analysis was carried out to determine whether the *MLH1* mutation in lbl-1260 is associated with aberrant expression of the protein. MLH1 expression was observed in both lbl-1261 and a control cell line, lbl-a, whereas expression was absent in a negative control cell line, HCT116 (Figure 3.2B). Intriguingly cell line lbl-1260 also demonstrated loss of MLH1 protein (Figure 3.2B). These data are surprising since, the germline *MLH1* mutation in lbl-1260 induces a single amino acid deletion (Hamilton *et al.*, 1995). Nonetheless, the extreme reduction of MLH1 in lbl-1260 suggests that the defects in MMR activity observed previously on a biochemical level, may not be due to a dominant negative effect as speculated (Parsons *et al.*, 1995a). Instead the profound loss of MLH1 protein likely confers the MMR deficiency in this cell line.

These data support the previous evidence that demonstrates cell lines lbl-1260 and lbl-1261 have defects in MMR (Parsons *et al.*, 1995a; Hamilton *et al.*, 1995; Nicolaides *et al.*, 1998). Furthermore, they shed new light on the mechanism by which the MMR defects in lbl-1260 arise.

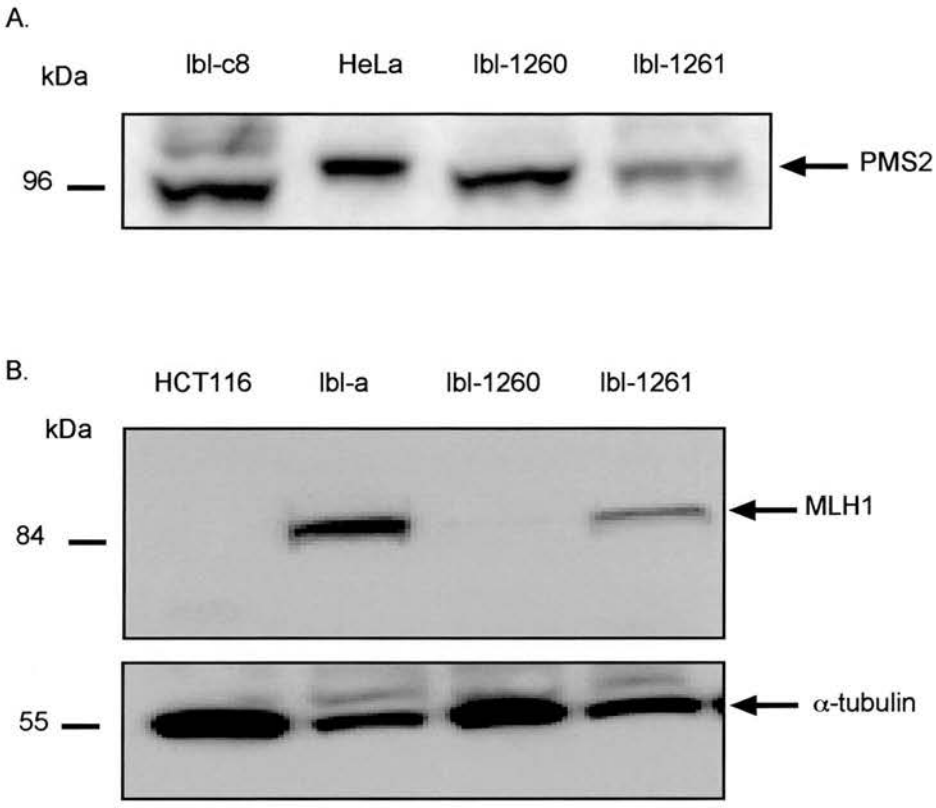


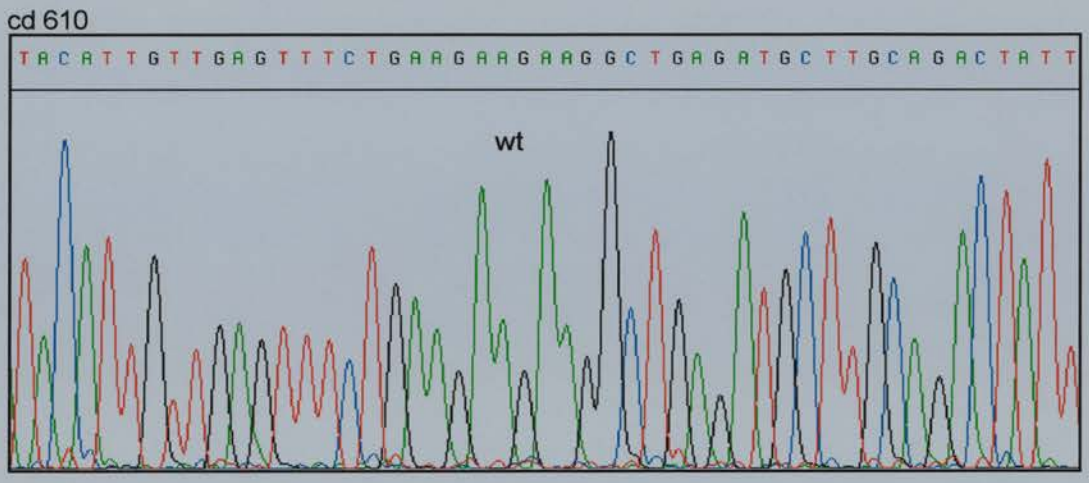
Figure 3.2 Western analysis of PMS2 and MLH1 expression in lbl-1261 and lbl-1260. (A) Wild type PMS2 expression is reduced by half in cell line lbl-1261 compared with positive control cell lines HeLa and lbl-c8 and MMR deficient cell line lbl-1260. The HeLa nuclear extract was comercially prepared and this likely explains why the HeLa PMS2 protein runs slightly slower than those from extracts prepared in this study. Equal concentrations of protein were loaded in each lane as determined by Bradford Assay. (B) lbl-1260 shows a severe reduction in MLH1 compared to MMR proficient control cell line lbl-a and MMR deficient cell line lbl-1261. MMR deficient cancer cell line, HCT116 provides a negative control for MLH1 expression. α-tubulin expression is used as a loading control.

3.3.3 Expression of mutant and wild type MLH1 mRNA transcripts in lbl-1260

The observation that lbl-1260 exhibits a severe reduction of MLH1, is particularly intriguing. It may be speculated that the *MLH1* mutation in lbl-1260 confers aberrant expression of the mutant allele. Loss of the wild type protein may result from a genomic deletion or an unidentified abrogation such as silencing of gene expression by methylation of the *MLH1* promoter (Herman *et al.*, 1998; Veigl *et al.*, 1998). Since sequencing of *MLH1* from the constitutional DNA of lbl-1260 around the 618 mutation, revealed mutant and wild type alleles, deletion of the wild type allele at the genomic level cannot explain the results from the western analysis.

The hypothesis that loss of MLH1 protein in lbl-1260 may be due to gene silencing was then addressed. RNA was extracted from lbl-1260 and subjected to RT-PCR, amplifying a 335bp product containing the site of the 618 mutation that was subsequently cloned. Sequencing of individual alleles from 6 clones revealed 3 mutant and 3 wild type alleles (Figure 3.3). The data suggests both *MLH1* alleles are transcribed in lbl-1260. Although a small number of transcripts were analysed, there is no evidence to indicate over or under expression of either the mutant or wild type mRNA. Since both mutant and wild type *MLH1* alleles are expressed at the mRNA level, the profound defect in MLH1 protein expression in lbl-1260 appears unlikely to be due to gene silencing.

A.



B.

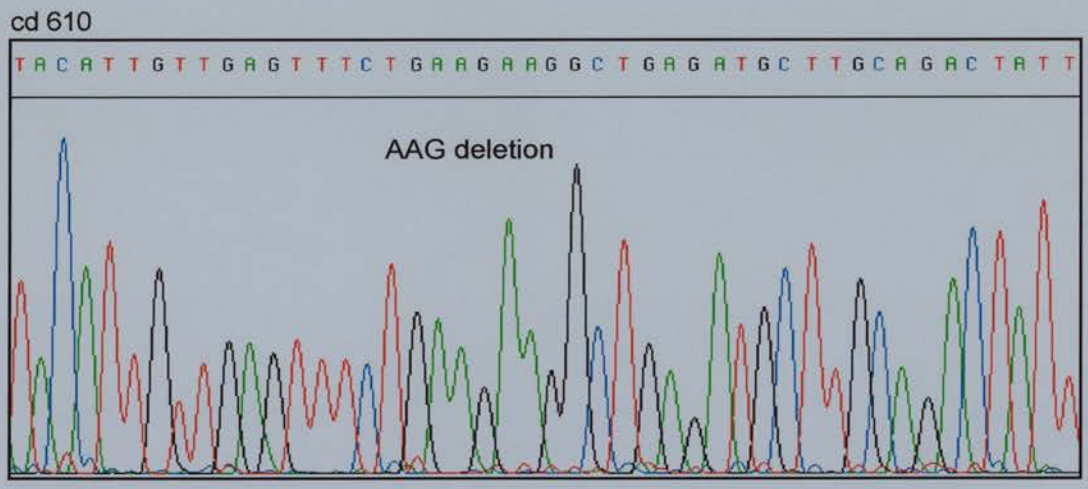


Figure 3.3 Representative sequences of individual *MLH1* transcripts surrounding the mutation at codon 618, cloned from lbl-1260 RNA by RT-PCR. (A) Clone 29 is wild type at codon 618 representing transcription of the wild type allele. (B) Clone 49 harbours the germline 618 mutation resulting in a 3bp AAG deletion and is indicitive of transcription of the mutant allele.

3.4 Discussion

Cell lines lbl-1260 and lbl-1261 were previously derived from the normal tissue of two patients with CRC, a family history of the disease and phenotypic evidence of TS (Hamilton *et al.*, 1995). Lbl-1260 and lbl-1261 harbour germline mutations in the MMR genes *MLH1* and *PMS2* respectively, and have been demonstrated to have profound defects in MMR on a biochemical level (Parsons *et al.*, 1995a). In this chapter, evidence has been presented that confirms the "normal" lineage of origin of these cell lines. Furthermore, the defects observed in the expression of the MLH1 and PMS2 proteins in lbl-1260 and lbl-1261 respectively, are in line with the constitutive defects in MMR that have been previously reported (Parsons *et al.*, 1995a).

Since these cells are utilised in subsequent experiments specifically because they are derived from non-cancerous tissue, demonstration of their lineage of origin is clearly valid. By using a panel of cell surface markers including those associated with B-cells, epithelial cells and a number of negative controls, it has been demonstrated that lbl-1260 and lbl-1261 express surface proteins consistent with them being of B-cell lymphocytic origin. In addition, there is no evidence that they express epithelial cell surface proteins indicative of a primary cancerous phenotype or micro-metastasis. Therefore, it is reasonable to conclude that lbl-1260 and lbl-1261 are derived from normal, non-cancerous tissue.

The cause of the MMR deficiency in lbl-1261 has been determined previously to be due to a dominant negative effect from the mutant PMS2 protein (Nicolaides *et al.*, 1998). The mutation results in a severe truncation of PMS2 and this is reflected by the reduced expression of the wild type protein detected in the western analysis here.

Interestingly, lbl-1260 does not express MLH1 to detectable levels. The data strongly argues against the suggestion that the lbl-1260 *MLH1* mutation encodes a protein that confers a dominant negative effect (Parsons *et al.*, 1995a). Rather, the MMR defects are likely due to the loss of expression of both wild type and mutant MLH1 proteins. Analysis of the *MLH1* mRNA species from lbl-1260 by RT-PCR

analysis revealed the presence of mutant and wild type transcripts. This analysis suggests that gene silencing of *MLH1* by promoter hypermethylation is unlikely to be the mechanism by which MLH1 protein expression is lost in lbl-1260. Promoter methylation inactivates the *MLH1* gene in the majority of MSI^+ sporadic CRCs (Cunningham *et al.*, 1998; Veigl *et al.*, 1998; Jones and Laird, 1999). However, *MLH1* silencing by methylation of the promoter region is correlated with decreased expression of the mRNA transcript in addition to loss of the protein (Herman *et al.*, 1998).

It is possible that the *MLH1* mutation in lbl-1260 affects the stability or translation of the transcript and this could be tested in vitro. It should also be noted that the region of binding, of the MLH1 antibody used in this study, has not been determined. If the antibody specifically recognises the region of the MLH1 protein corresponding to the 618 mutation, the mutant protein may not be detected. However, this does not explain why the wild type protein is not detected. Further investigation is clearly required and will be of interest in order to establish the nature of the loss of MLH1 protein in lbl-1260. Unravelling the exact cause may reveal a novel mechanism of MMR inactivation.

The experiments presented in this chapter demonstrate conclusively that lbl-1260 and lbl-1261 are derived from normal tissue and the western analyses of protein expression is entirely consistent with MMR defects previously reported previously (Hamilton *et al.*, 1995; Nicolaides *et al.*, 1998).

In the following chapters cell lines lbl-1260 and lbl-1261 have been utilised to effectively unmask specific hypermutable loci, at both non-coding and coding sequence in view of the fact that the frequency of mutations will not be biased by other factors involved in tumourigenesis.

Chapter 4

MMR Defects in Cells Derived from Normal Tissue Are Associated with Microsatellite Instability

4.1 Introduction

Simple sequence repeats (SSRs) account for up to half of the human genome (Toth *et al.*, 2000). Many are highly polymorphic, making them of particular importance to the study of evolution and the mapping of disease genes (Weissenbach *et al.*, 1992; Weber and Wong, 1993). New mutations occur at such sequences predominantly as a result of slippage during replication (Levinson and Gutman, 1987) and failure to correct these errors in the presence of MMR defects results in MSI (Ionov *et al.*, 1993; Strand *et al.*, 1993; Thibodeau *et al.*, 1993). In view of the susceptibility of microsatellites to mutation accumulation in the presence of MMR defects, both poly(A/T) and (CA)n microsatellites are routinely employed to determine the MSI status of colorectal tumours (Dietmaier *et al.*, 1997; Rodriguez-Bigas *et al.*, 1997; Boland *et al.*, 1998).

Differences in mutation rate are evident at alternative microsatellite loci both in populations (Weissenbach *et al.*, 1992) and in MMR deficient tumours (Thibodeau *et al.*, 1993; Shibata *et al.*, 1994; Hoff-Olsen *et al.*, 1998; Thibodeau *et al.*, 1998; Zhang *et al.*, 2001). Variation in the repeat sequence itself can significantly affect individual microsatellite stability (Weber and Wong, 1993; Bull *et al.*, 1999). Mononucleotide sequences are particularly prone to mutations (Chen *et al.*, 1995) but even these can display significant variation in mutation frequency dependent in part on the nature of the mononucleotide tract (A/T vs G/C) (Zhang *et al.*, 2001). In addition, variations in microsatellite flanking sequence (Bichara *et al.*, 1995; Andreassen *et al.*, 1996; Zhang *et al.*, 2001), and in repeat length (Wierdl *et al.*, 1997; Xu *et al.*, 2000) have been demonstrated to contribute to heterogeneity in mutation rate.

The use of MMR deficient cancer cells to investigate inherent stability of microsatellites, and factors that may influence the manifestation of the mutator phenotype, suffers from several limitations. Mutational homogeneity can arise at microsatellites in cancer cells due to clonal evolution of the tumour (Nowell, 1976; Shibata *et al.*, 1994; Cahill *et al.*, 1999). Additional mutations may accumulate during expansion of the tumour cell population, but further waves of clonal selection acting on advantageous coding sequence mutations, will again reduce the mutational variation of non-transcribed microsatellite sequences. Consequently, the spectrum and frequency of microsatellite mutations will be biased, due to the effects of selection and other variability's that are dependent on the process of tumourigenesis. This makes it difficult to identify mutational events that occur exclusively as a result of MMR defects in cancer cells.

Data presented in Chapter 3 and elsewhere, has demonstrated that the germline mutations in cell lines lbl-1260 and lbl-1261 result in complete loss of MMR activity (Parsons *et al.*, 1995a). In view of the fact that MSI is an excellent indicator of MMR defects, these cell lines derived from normal tissue, were analysed by a small pool PCR (SP-PCR) approach at two microsatellite repeat loci to address whether the repair defects in lbl-1260 and lbl-1261, are associated with a mutator phenotype. Microsatellite analysis was carried out at markers, D2S123 and BAT-40, in MMR deficient lymphoblast cell lines lbl-1261 and lbl-1260 as well as MMR proficient control lymphoblast cell lines lbl-c1 and lbl-c5. D2S123 is a (CA)n microsatellite marker (GenBank accession number; Z16551) and BAT-40 is a poly(A) marker (GenBank accession number; M38180). D2S123 is one of the panel of five markers, recommended for use in determination of microsatellite status of colorectal tumours (Rodriguez-Bigas *et al.*, 1997; Boland *et al.*, 1998). BAT-40 is not one of the panel of five, but is recommended for use in MSI analysis as part of an expanded panel of markers (Boland *et al.*, 1998).

It was anticipated that mutations in these cell lines would arise at a low frequency within a population of predominantly wild type cells. Hence a sensitive small pool SP-PCR strategy was adapted to genotype individual microsatellite alleles from these cell lines. A similar approach has been used previously to detect mutations within

populations of wild type alleles (Jeffreys *et al.*, 1994; Monckton *et al.*, 1995; Parsons *et al.*, 1995a; Vilkki *et al.*, 2001)

This sensitised system was also utilised to investigate whether factors independent of tumourigenesis, can influence stability at poly(A) and (CA) repeats when MMR is absent. Identifying determinants of instability, which may affect the manifestation of a mutator phenotype at repetitive sequences, is clearly merited in view of the relevance to analysis of tumour MSI status. Since cell lines lbl-1260 and lbl-1261 are derived from normal cells, the full spectrum of mutations arising consequent of MMR defects was predicted to be available for analysis. This overcomes problems associated with clonal evolution of microsatellite sequences in cancer cells and mutational bias from tumourigenic process itself.

4.2 Methodological Overview

4.2.1 DNA extraction

DNA was extracted from patient cell lines lbl-1260 and lbl-1261 and control lymphoblast cell lines lbl-c1 and lbl-c5 as described in 2.2.1. DNA concentration was determined as described in 2.2.5.

4.2.2 Small pool PCR

Limiting dilution experiments and SP-PCR was carried out as described previously in 2.4.4 and 2.4.5 using primers BAT-40F and BAT-40R and D2S123F and D2S123R (2.4.1). BAT-40 forward primers were fluorescently labelled with HEX and D2S123 forward primers were fluorescently labelled with FAM. High Fidelity Taq polymerase (Expand, Boehringer Mannheim) was used in all PCRs. This has a three fold increased fidelity of DNA synthesis compared to Amplitaq (PE Applied Biosystems) to achieve the most faithful replication of microsatellite alleles. Since amplifications involved very dilute DNA, rigorous controls were employed against contamination and these are detailed in 2.4.5 and 2.5.1.

4.2.3 Mutational analysis of SP-PCR products

SP-PCR products were size analysed using an ABI310 genetic analyser as described in 2.5.1 and the size of each allele detected was recorded.

4.2.4 Data analysis

The frequency of mutant alleles in each cell line was expressed as the number of alleles that were mutant in length divided by the total number of alleles detected (normal and mutant). Accordingly, percentages are not exact contents of cells with

alterations, but relative values of alleles. Differences in the frequency of normal and mutant alleles between cell lines were evaluated by a χ^2 test as described in 2.10.1 and significance taken at 5%.

Where mutant allele origin was assigned as having derived from one of two wild type constitutional alleles (relating to analysis of D2S123 mutant profiles in cell line lbl-1261), this was determined as described in 2.5.4. To subsequently evaluate whether there was a bias in mutation at either wild type allele, this was carried out using a χ^2 test to compare the number of mutant and wild type alleles at each progenitor (2.10.1). To account for the possibility that some mutants did not derive from the progenitor allele closest in size, a further more stringent assessment was also performed and this is detailed in the results.

4.3 Results

The results in this chapter have also been published in Bacon *et al.*, 2000 and in Bacon *et al.*, 2001b.

4.3.1 Detection of mutant microsatellite alleles in lbl-1260 and lbl-1261

Genotyping of D2S123 alleles from MMR defective cell lines lbl-1261 and lbl-1260 revealed constitutional allele sizes of 214/228bp (Figure 4.1) and 212/228bp respectively. Genotyping of D2S123 alleles from MMR proficient cell line lbl-c5 revealed alleles of 212/214bp. At a second poly(A) marker, BAT-40, genotyping of alleles from lbl-1261 and lbl-1260 revealed constitutional alleles of 120/124bp (Figure 4.2) and 124/124bp respectively. At two-control cell lines, lbl-c5 and lbl-c1 constitutional allele sizes were 120/124bp and 120/124bp respectively.

Constitutional allele sizes as determined from the amplification of undiluted DNA were subsequently confirmed during SP-PCR analysis. Individual alleles of the same size as those determined from undiluted DNA were detected, validating the method of genotyping used (Figures 4.1 and 4.2). In addition, reproducibility in the amplification and sizing technique was evidenced by consistent sizing of positive PCR samples from two wells on every SP-PCR plate.

For SP-PCR analysis, the extreme dilute nature of the DNA was evidenced in all analyses by the absence of any peaks in the majority of PCRs or, the detection of single alleles from heterozygous samples in others (Figures 4.1 and 4.3). In all, between 54 and 270 alleles were typed by SP-PCR analysis in each of the cell lines at both markers. Length variation in individual alleles was clearly detected at D2S123 (Figure 4.3) and BAT-40 (Figure 4.4) in both MMR deficient cell lines, by comparison to the constitutional allele sizes in undiluted DNA.

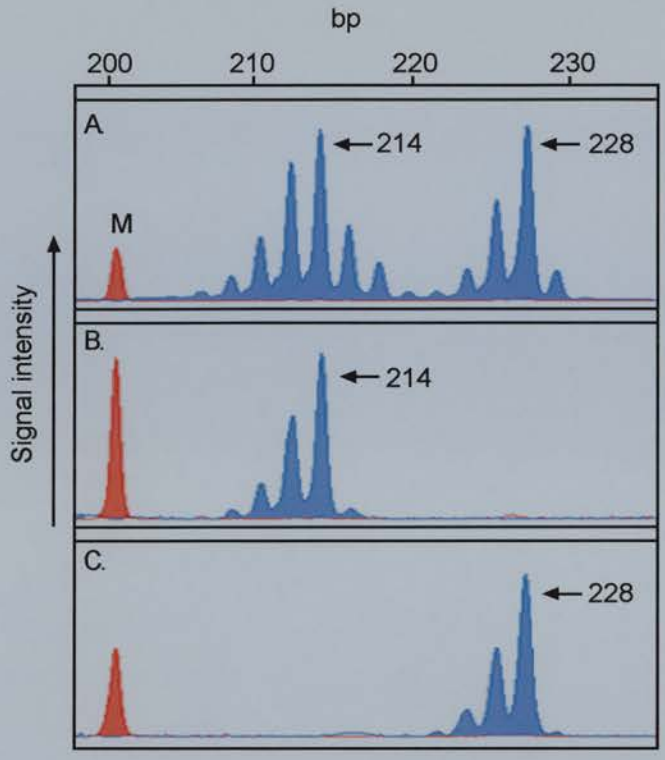


Figure 4.1 (A) Constitutional genotype of DNA from Ibl-1261 at D2S123. Genotype is determined from the preparation of undiluted DNA. A short and a long allele of 214 and 228bp respectively are clearly detected as the predominant peaks in each peak complex. (B and C) This is confirmed by genotyping individual alleles from Ibl-1261 using SP-PCR where alleles of 214 and 228bp are those most frequently detected. M = marker peak at 200bp.

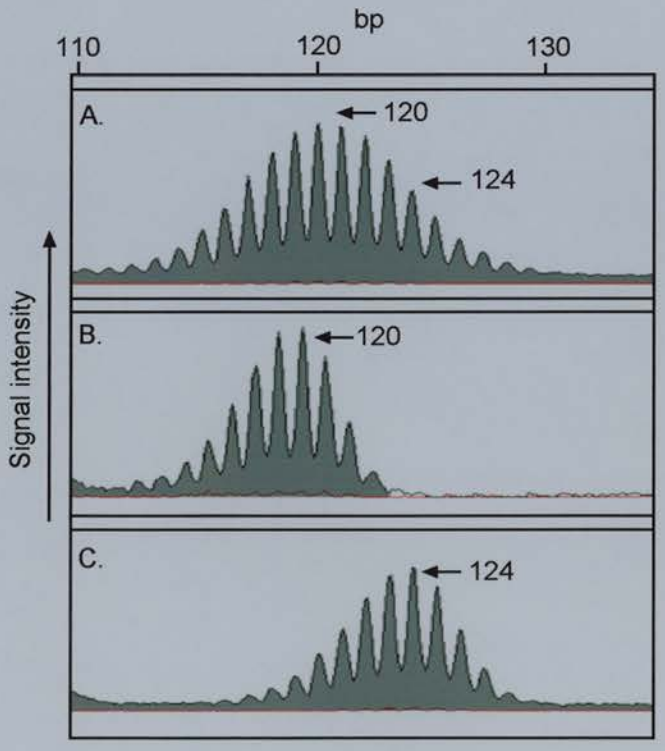


Figure 4.2 (A) Constitutional genotype of DNA from Ibl-1261 at BAT-40 determined from the preparation of undiluted DNA. Wild type allele sizes are taken as the predominant peak in each peak complex as determined from genescan data. (B and C) the constitutional allele sizes of 120 and 124bp are confirmed in SP-PCR analysis. Individual alleles of 120 and 124bp are easily determined and are the most frequently observed allele sizes.

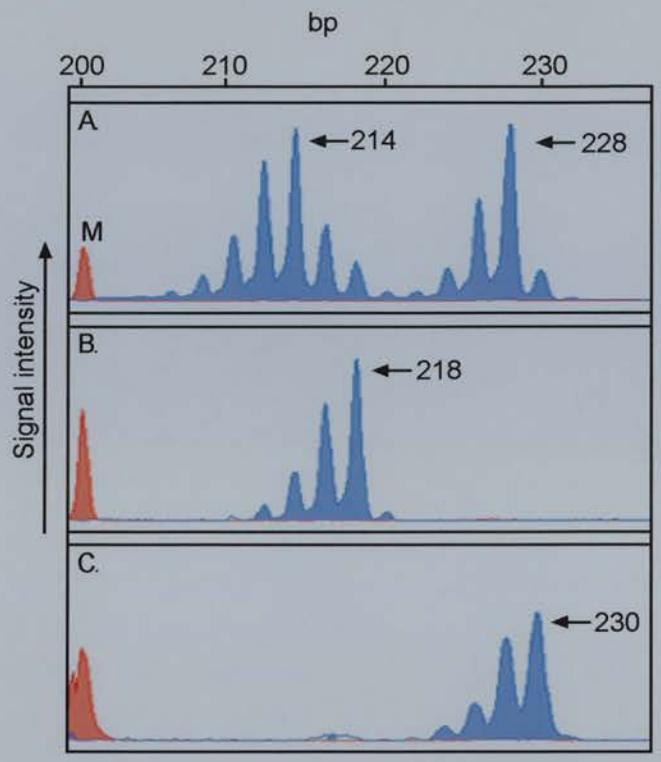


Figure 4.3 Detection of mutant D2S123 alleles in Ibl-1261 by Sp-PCR. (A) Constitutional genotype in undiluted DNA. (B and C) In SP-PCR analysis individual mutant alleles of 218 and 230bp are clearly detected in Ibl-1261 by comparison to constitutional genotype. M = marker peak at 200bp.

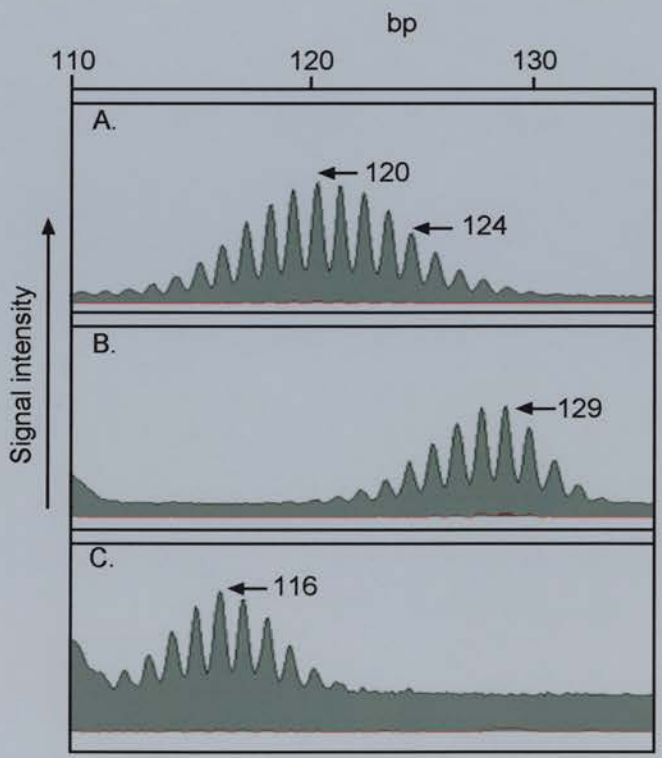


Figure 4.4 (B and C) Detection of mutant BAT-40 alleles in Ibl-1261 by SP-PCR. Individual mutant alleles of 129 and 116bp are detected in Ibl-1261 by SP-PCR. (A) The mutant alleles are clearly identified by comparison to the constitutional genotype as determined from undiluted DNA (A).

4.3.2 Cells derived from normal tissue with MMR defects display a mutator phenotype

The size profiles and frequency of individual D2S123 and BAT-40 alleles typed from each cell line by SP-PCR are shown in Figures 4.5 and 4.6 and the frequency of mutants detected in each cell line is summarised in Table 4.1.

Table 4.1 Summary of SP-PCR analysis at microsatellite markers. Mutant allele counts at D2S123 and BAT-40 SSR loci in MMR deficient lymphoblastoid cell lines lbl-1261, lbl-1260 and MMR proficient lymphoblastoid cell lines, lbl-c5, lbl-c1 are shown. MMR deficient cell lines are significantly more unstable than control cell lines (χ^2=84.032, p<0.001 MMR deficient vs proficient cell lines at both markers.

Cell line	D2S123		BAT-40	
	Total alleles	Mutants (Frequency)	Total alleles	Mutants (Frequency)
lbl-1261	270	89 (0.330)	139	71 (0.51)
lbl-1260	120	10 (0.083)	54	13 (0.24)
lbl-c5	109	2 (0.018)	55	1 (0.02)
lbl-c1	ND	ND	99	5 (0.05)

Both MMR defective cell lines displayed elevated levels of mutation frequency compared to MMR proficient cell lines at D2S123 (Figure 4.5) and BAT-40 (Figure 4.6). 89/270 (33%) D2S123 alleles and 71/139 (51.1%) BAT-40 alleles typed from lbl-1261 were mutant as were 10/120 (8.3%) and 13/54 (24.1%) alleles typed from lbl-1260. These mutation frequencies were significantly increased relative to the control cell lines both at D2S123 (χ^2=41.2, p<0.001; χ^2=4.9, p=0.028 for lbl-1261 and lbl-1260 respectively) and BAT-40 (χ^2=85.02, p<0.001; χ^2 =19.9; p<0.001 for lbl-1261 and lbl-1260 respectively)

The low number of mutants observed in control cell lines may be due to PCR error or indeed could be genuine mutations. However their small number relative to those in MMR deficient cell lines indicates that PCR artefacts do not interfere significantly with the detection of mutants in the dilute DNA.

Heterogeneity was observed in the level of mutation between the two MMR deficient cell lines. Lbl-1261 displayed a significantly higher level of mutation compared to lbl-1260 at both BAT-40 and D2S123 (χ^2=38.0, p<0.001; markers combined). In addition clustering of mutants around wild type alleles was observed suggesting the predominance of small mutations. A wide spectrum of insertion and deletion mutant alleles was observed at both marker loci indicating that PCR bias towards shorter alleles is not a significant factor of the technique used.

These results indicate that cell lines lbl-1261 and lbl-1260 that are derived from normal tissue display a mutator phenotype similar to that observed in MMR deficient tumours. These data demonstrate that the MMR defects in these cell lines are associated with microsatellite instability and that when selection pressures and confounding changes associated with cancer development are minimised, a wide spectrum of mutations consequent of repair defects can be detected.

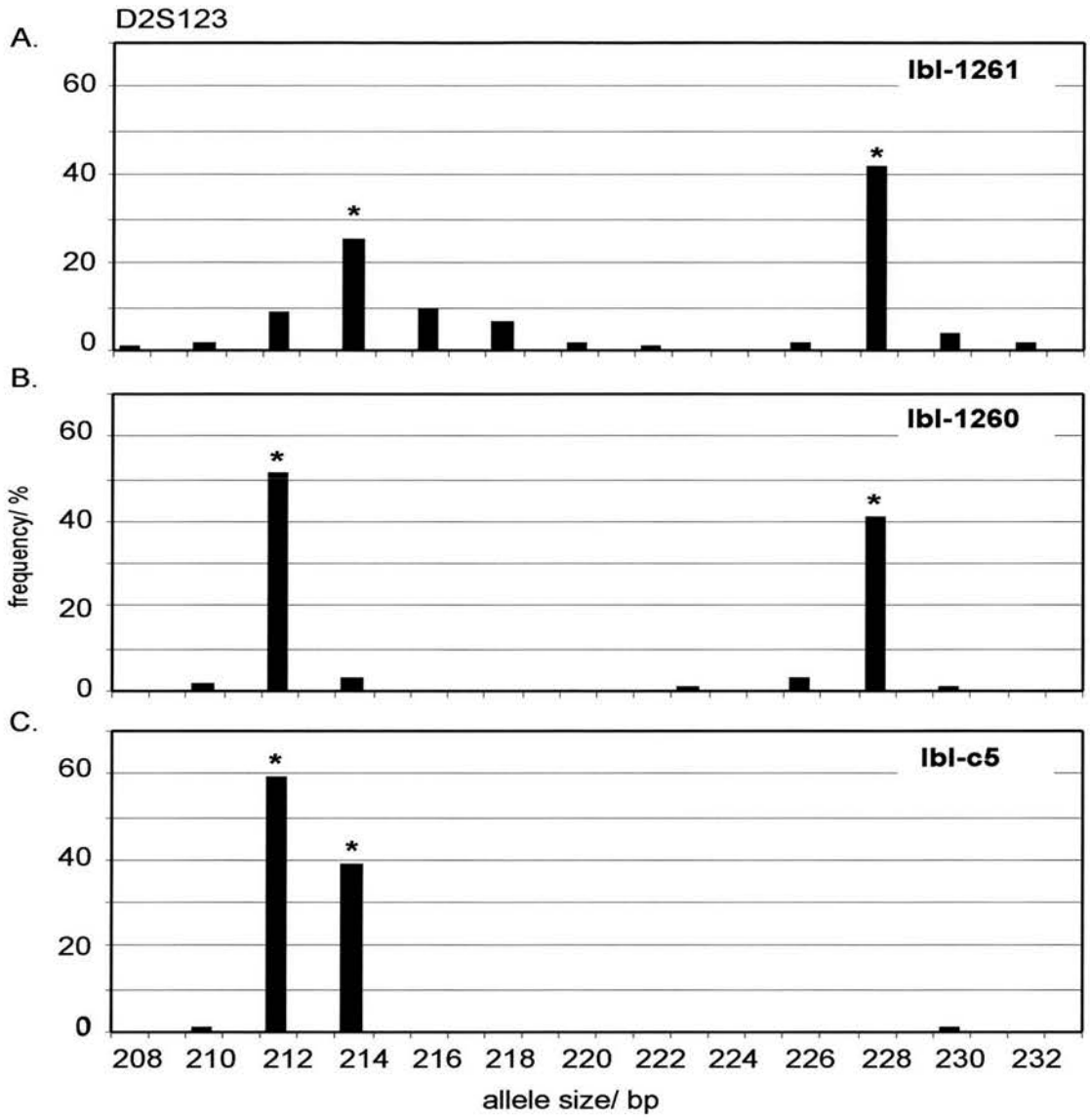


Figure 4.5 Microsatellite Instability at D2S123 in non-cancer derived cells with MMR defects. Distribution of D2S123 allele sizes in MMR deficient (A and B) and MMR proficient (C) lymphoblastoid cell lines as detected by SP-PCR. Asterisks indicate wild type alleles as identified by analysis of undiluted DNA. A significantly elevated frequency of mutation is demonstrated in MMR deficient cell lines lbl-1260 and lbl-1261, compared to control cell line lbl-c5 ($p<0.03$, lbl-1260 and lbl-1261 data combined).

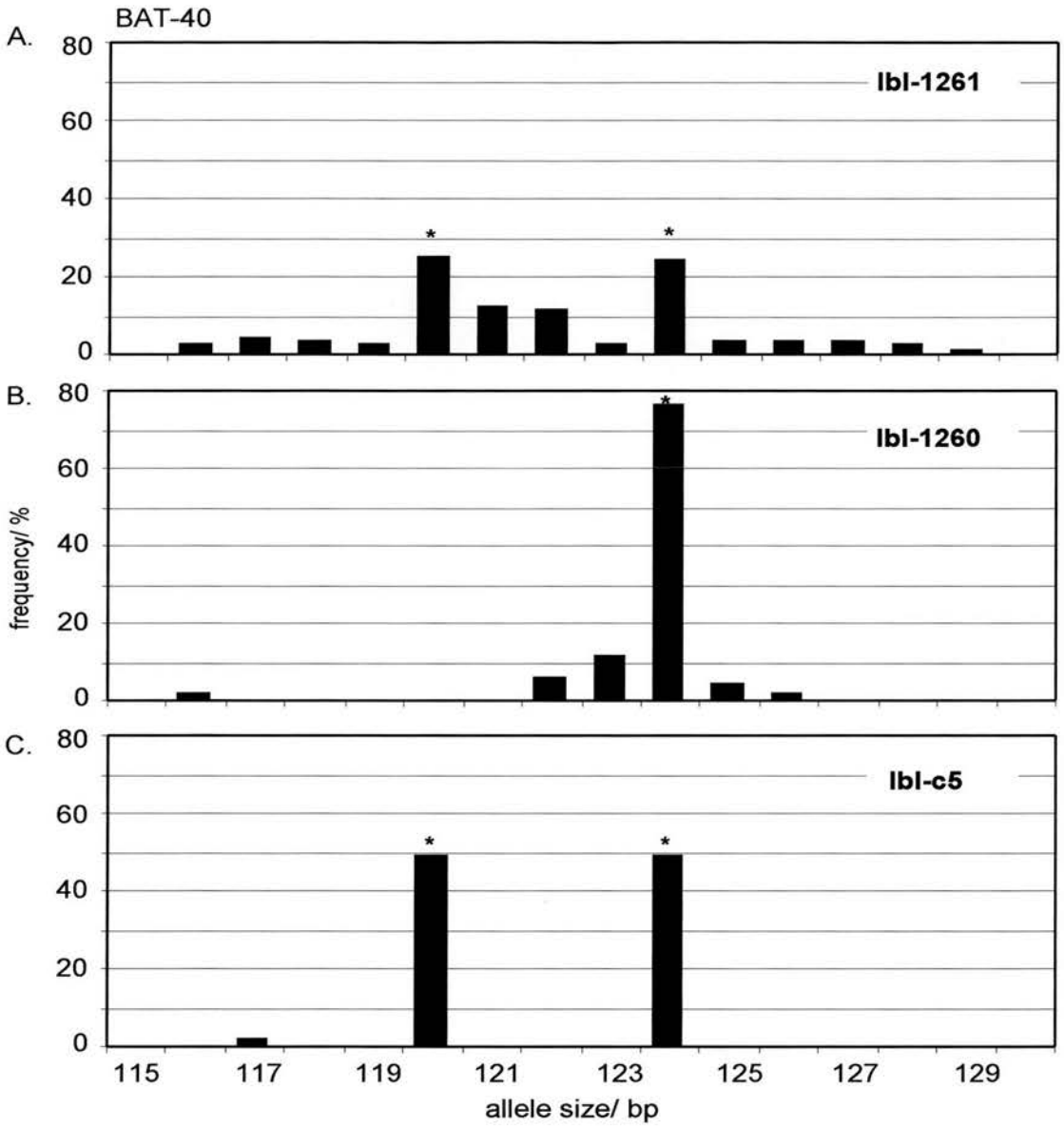


Figure 4.6 Microsatellite Instability at BAT-40 in non-cancer derived cells with MMR defects. Distribution of BAT-40 allele sizes in MMR deficient (A and B) and MMR proficient (C) lymphoblastoid cell lines as detected by SP-PCR. Asterisks indicate the predominant constitutional alleles identified by analysis of undiluted DNA. Significant elevation in the frequency of mutations is demonstrated in MMR deficient cell lines compared to MMR proficient controls ($p<0.001$, lbl-1260 and lbl-1261 data combined).

4.3.3 Allele bias in mutation at D2S123 in MMR deficient cells

It was noticed that there was a substantial bias in the wild type constitutional alleles from which mutants were derived, at microsatellite D2S123. This was exacerbated in the most unstable cell line, lbl-1261. In all, 71 of 89 mutants (80%) in lbl-1261 were clustered around the shorter (214bp) allele (Figure 4.5), indicating a substantial bias for mutation at that locus (χ^2=42.6, p<0.001). Mutant alleles were assumed to be derived from the progenitor allele in the undiluted DNA that was closest in size since, studies of microsatellite mutations in human pedigrees, human cell lines and artificial constructs in yeast, have shown that mutations in most microsatellites involve only one or two repeat units (Weber and Wong, 1993; Petes *et al.*, 1997; Olsen *et al.*, 1998). Since mutant alleles in lbl-1261 ranged in size from 208-232bp and the progenitor alleles were 214 and 228bp, mutants of size ≤220bp were assigned as derivatives from the 214bp-progenitor allele. Mutants of ≥222bp were assigned as derivatives from the 228bp-progenitor allele. However, to ensure that the observed bias was not due to mis-assignment of mutant alleles, the data was re-analysed using a threshold to exclude large reductions in the 228bp allele. It was assumed that only mutants of ≤216bp (a single repeat expansion of the short progenitor allele) were derived from the shorter progenitor. This confirmed the highly significant mutation bias at the shorter allele (χ^2=9.43, p=0.002).

These observations are not due to PCR bias at the shorter allele, since equal numbers of short and long alleles were detected in lbl-1261. Furthermore, insertion and deletion mutants of both progenitor alleles were detected (Figure 4.5). These observations indicate a substantial and statistically significant allele-specific bias in mutation rate at D2S123. These observations are investigated further and discussed in Chapter 5 where the nature of this mutational bias is determined.

4.3.4 Evidence of increased instability at microsatellite BAT-40

Interestingly, instability was heterogeneous between the two microsatellites D2S123 and BAT-40 (Table 4.1). The level of mutation at BAT-40 was significantly elevated in comparison to D2S123 in both cell lines ($\chi 2=19.7$, $p<0.001$) suggesting differential susceptibility of SSRs for mutation. The increased frequency of mutation at BAT-40 indicates the poly(A) tract has a significantly greater propensity to mutation than the (CA)n repeat in D2S123, and may be inherently very unstable. Further investigation to determine if BAT-40 is a hypermutable locus was carried out and this is detailed and further discussed in Chapter 6.

4.4 Discussion

In this chapter, data has been presented that demonstrates a substantial excess of mutant microsatellite alleles can be detected in cell lines derived from normal tissue with MMR defects.

Profound MMR defects in lbl-1261 and lbl-1260 have been demonstrated previously (Parsons *et al.*, 1995a; Nicolaides *et al.*, 1998) and the levels of MMR protein expression in these cell lines are consistent with such defects (Chapter 3). Analysis of individual alleles by SP-PCR has shown that these defects in MMR are associated with low level microsatellite instability in these cell lines, confirming a previous report (Parsons *et al.*, 1995a). Furthermore, this investigation demonstrates that the mutator phenotype displays significant heterogeneity with regards to both the spectrum and frequency with which mutations occur at microsatellite sequences and suggests that D2S123 and BAT-40 have differential inherent susceptibilities to mutation.

These data provide further support for a growing number of studies that show constitutional MMR defects can result in widespread microsatellite instability in normal tissue. High levels of DNA instability in non-cancerous tissues has been demonstrated in cases of knockout mice with homologous mutations of *Pms2* and *Msh2* genes (Baker *et al.*, 1995; de Wind *et al.*, 1995; Narayanan *et al.*, 1997). In addition rare individuals have been identified with biallelic germline mutations in the *MLH1* gene (Ricciardone *et al.*, 1999; Wang *et al.*, 1999; Vilkki *et al.*, 2001) and these also show instability at microsatellite loci in normal tissue. It is interesting to note that the widespread MSI in the non-cancerous tissue of these patients is associated with early malignancies, including extra-colonic cancers (Ricciardone *et al.*, 1999; Wang *et al.*, 1999; Vilkki *et al.*, 2001). The cell lines analysed in this study were also derived from the normal tissue (Chapter 3) of patients, with early onset cancer and who developed extra-colonic lesions displaying phenotypic evidence of TS (Hamilton *et al.*, 1995). In a separate study, a Turcots patient with a germline *PMS2* mutation also showed instability in normal tissue (Miyaki *et al.*, 1997b). Taken together these studies suggest that constitutive loss of MMR results in

widespread MSI and this likely contributes to severe forms of early onset malignancies. It will be of interest to establish how widespread this phenomenon may be.

Insertion and deletion mutations were observed at both markers, with small changes predominating. This is in agreement with previous studies in both yeast and MMR deficient tumours demonstrating that most microsatellite alterations involve the addition or subtraction of one or two repeating units (Brinkmann *et al.*, 1998; Sturzeneker *et al.*, 2000). However a wide spectrum of mutations was also observed, most dramatically in cell line lbl-1261. This suggests that mutational bias due to selection and clonal expansion in these cell lines is not a significant factor.

The observed variation in the level of mutation between cell lines lbl-1260 and lbl-1261, likely reflects the different mutations that they harbour. It is well documented that defects in various MMR genes result in different levels of mutation accumulation. *MSH6* germline mutations are known to result predominantly in a microsatellite instability low (MSI-L) phenotype due to redundancy between the MSH6 and MSH3 proteins (Marsischky *et al.*, 1996; Wu *et al.*, 1999b). However, the majority of tumours with high levels of microsatellite instability (MSI-H) are accounted for by mutations of the *MLH1* and *MSH2* genes (Liu *et al.*, 1996; Thibodeau *et al.*, 1998). In a study of Mlh1 versus Pms2 deficient mice, the mononucleotide repeat mutation frequency was consistently 2- to 3- fold higher in the $Mlh1^{-/-}$ compared to the $Pms2^{-/-}$ mice (Yao *et al.*, 1999). Interestingly, this is in contrast to the results presented here, in which a 3- fold increase in mutation was observed in lbl-1261 (*PMS2* mutation) compared to lbl-1260 (*MLH1* mutation). However, cell lines with mutations in the same gene have also been demonstrated to show significant differences in the extent of mutation at dinucleotide repeats (Oki *et al.*, 1999), indicating the nature of the MMR mutation, as well as the specific gene, influences the manifestation of the mutator phenotype. The increased frequency of mutation observed in cell line lbl-1261 is also in agreement with a previous study (Parsons *et al.*, 1995a).

The statistically significant allele bias for mutation observed at D2S123 in cell line lbl-1261, suggests that features of the sequence or chromatin structure at the

shorter progenitor allele may contribute to the increased susceptibility for mutation. Alternatively, some influence at the long progenitor allele may confer stability. Intra-allelic heterogeneity for mutation adds further complexity to understanding the manifestation of the mutator phenotype. That the locus genotype might determine susceptibility to mutation, is of relevance when evaluating potential targets of the mutator phenotype as well as in the use of such markers for the classification of MSI status of colorectal tumours. This observation is also of interest since it conflicts with the expectation that MSI correlates with increased repeat length and number of repeating units (Sia *et al.*, 1997; Wierdl *et al.*, 1997). The nature of this allele bias for mutation is further investigated in Chapter 5.

Heterogeneity in mutation frequency was observed between BAT-40 and D2S123 loci, with a statistically significant elevation in mutation at BAT-40 in both cell lines. BAT-40 showed an overall 1.7 fold increase in mutation frequency over D2S123. In lbl-1260, the level of BAT-40 mutation was 3 times that of D2S123. These data indicate differences also exist in the inherent stability at alternative loci and that BAT-40 is particularly prone to mutation. Previous studies have demonstrated that microsatellite repeat loci containing mononucleotide repeat motifs particularly of the type poly(A)n, are more susceptible to instability in MMR deficient tumours (Dietmaier *et al.*, 1997). Furthermore, BAT-40 has been reported to show mutation in over 95% of MSI cancers (Parsons *et al.*, 1995b). Extreme inherent instability at this locus may explain the increased sensitivity to mutation in the presence of MMR defects and this hypothesis is addressed in Chapter 6.

In summary, the data in this chapter indicates that defects in MMR in cells derived from normal tissue are associated with microsatellite instability of the type observed in MMR deficient tumours and are in agreement with a previous study (Parsons *et al.*, 1995a). This study illustrates that MMR deficient cell lines derived from normal tissue are excellent tools in unmasking underlying mutation frequencies and revealing factors that may affect the manifestation of a mutator phenotype or result in the accumulation of new mutations at microsatellite sequences.

The fact that heterogeneity in the mutator phenotype was detected on a number of different levels, suggests that the manifestation of MSI can be affected by a number

of different factors. Variation in the mutation frequency at a given microsatellite loci may be a result of the nature of the MMR mutation, or may be due to features of the sequence within or surrounding individual repetitive tracts in both an allele specific or locus specific manner. The contribution to mutability from inherent sequence characteristics are of particular relevance when employing microsatellite markers in the classification of MSI tumours and also in evaluating sequences that may be potentially unstable consequent of MMR defects. These initial findings are addressed in further detail in the next two chapters.

Chapter 5

Sequence Interruptions Confer Differential Stability at Microsatellite Alleles in MMR Deficient Cells

5.1 Introduction

In Chapter 4, SP-PCR analysis of cells with MMR defects, derived from normal tissue at a (CA)n repeat marker, D2S123, demonstrated a substantial and statistically significant allele-specific bias in mutation frequency. The shorter progenitor allele in cell line lbl-1261 was found to be significantly more unstable than the longer allele, conflicting with the expectation that instability correlates with repeat length (Sia *et al.*, 1997; Wierdl *et al.*, 1997). This suggests that inherent features of the sequence within or surrounding the (CA) repeat at individual D2S123 alleles may contribute to the differential propensity to mutability observed at this locus.

The degree of perfection of a repeat has been demonstrated to correlate with instability. In *E.coli* and yeast, presence of an interspersion within a dinucleotide tract results in its stabilisation (Bichara *et al.*, 1995; Petes *et al.*, 1997). The stabilising effect of repeat interspersion was also observed in yeast strains with MMR defects (Petes *et al.*, 1997). Locus by locus analyses reveal that interspersed repeat markers are also relatively more stable in the fruitfly *Drosophila* (Goldstein and Clark, 1995), in the germline of human populations (Brinkmann *et al.*, 1998) and in tumours from cancer patients (Sturzeneker *et al.*, 2000). Repeat sequence interruption is also important on a clinical level. Expansion of triplet repeats in the *FMR1* gene, the *HD* gene and the *SCA1* gene, gives rise to the human trinucleotide disorders fragile X, Huntingdon's disease and spinocerebella ataxia type 1 respectively (Ashley and Warren, 1995; Gordenin *et al.*, 1997). Stability of both *SCA1* and *FMR1* alleles are conferred by interruption of the contiguous repeat (Chung *et al.*, 1993; Eichler *et al.*, 1994; Hirst *et al.*, 1994; Snow *et al.*, 1994; Kunst *et al.*, 1997).

The majority of these studies demonstrate variation in mutation frequency between alternative loci. Mechanisms affecting mutation rate within a given locus have been less extensively characterised, although such variation has been documented. For example in a population study of meioses in pentameric and tetrameric markers, it was shown that different alleles at a single locus varied in their rate of mutation (Brinkmann *et al.*, 1998).

In this chapter the nature of the allele specific mutational bias at D2S123 observed in lbl-1261 (Chapter 4), was determined. The wider relevance of the initial observations identified in lbl-1261 is established, by analysis of population D2S123 genotypes and the underlying cause of this mutational bias has been elucidated. The question as to whether constitutional genotype can influence susceptibility to mutation at D2S123 in MSI^+ CRCs was then specifically addressed by analysis of matched normal and tumour DNA. As well as providing considerable insight into understanding the mechanisms that lead to instability at repetitive sequences, this series of experiments has potential clinical relevance. The CA repeat marker studied here is one of a panel of five markers recommended for use in MSI analysis in colorectal cancer (Boland *et al.*, 1998; Perucho, 1999). Therefore, establishing whether patient genotype can influence the manifestation of MSI at this marker is of considerable importance. In addition, uncovering factors that affect repeat stability in the absence of MMR is important when considering other sequences, both coding or non-coding, that may show instability in the presence of MMR defects.

5.2 Methodological overview

5.2.1 Biological samples

The size distribution of D2S123 alleles within the population was determined by genotyping blood DNA of 39 unrelated Scottish individuals described in 2.3.2. Genotyping of matched normal and tumour samples was carried out using DNA's previously isolated from 41 colorectal cancer patients shown to have defective MMR and a mutator phenotype as described in 2.3.3 and 2.3.4. All patients with suitable material whose tumours fulfilled MSI^+ criteria were analysed in this study, regardless of whether or not D2S123 showed length variation in tumour DNA. There is potential bias because MSI criteria include D2S123 genotyping and so some patients with homozygous long alleles may be underscored with respect to MSI status. However this does not impact adversely on the finding of these investigations.

5.2.2 Genotyping samples at locus D2S123

For genotyping of DNA samples at the D2S123 locus, DNA templates were amplified as described in 2.4.2 using FAM labelled fluorescent D2S123F and D2S123R primers (2.4.1). The Expand High Fidelity PCR system (Boehringer Mannheim) was used, to ensure faithful replication over the repetitive tract. Size analysis of PCR products was carried out using an ABI310 genetic analyser as described in 2.5.1. with allele size taken to be the most predominant peak in each peak complex.

5.2.3 Sequencing of D2S123 alleles

D2S123 alleles were amplified from undiluted cell line or blood DNA as described in 2.4.2. using non-fluorescent D2S123F and D2S123R primers (2.4.1). Again, the Expand High Fidelity PCR system (Boehringer Mannheim) was used. Individual progenitor alleles were visualised by gel electrophoresis as described in

2.4.6 and gel purified as described in 2.7.1. Sequencing of purified products was then performed as described in 2.7.2-2.7.4.

5.2.4 Cloning of individual D2S123 alleles

D2S123 alleles from lbl-1261 and lbl-c5 were cloned as described in 2.6.1 and 2.6.2. Genotyping and cycle sequencing of transformants was performed as described in 2.5.1 and 2.7.1-2.7.5.

5.2.5 Determination of allele specific bias in MSI$^+$ tumours

The D2S123 locus was amplified from 100ng of normal and tumour DNA templates in 50µl PCR reactions with FAM labelled D2S123F and D2S123R primers as described in 2.4.1 and 2.4.2. PCR products were size analysed using an ABI 310 genetic analyser as described in 2.5.1. Analysis of individual shifts was made by comparison of normal and tumour Genescan profiles. Whenever there were doubts about the veracity of a mutation, it was not included. On some occasions it was questionable as to whether one or both alleles had mutated. Such cases were scored as shifting at just one of the alleles. Mutation frequency may therefore be marginally underestimated if both alleles had mutated and the presence of any wild type allele in the tumour represented contamination from surrounding normal mucosa. Observed differences in mutation frequency at each allele were evaluated by χ^2 analysis as described in 2.10.1

5.3 Results

The results presented here have been published in Bacon *et al.*, 2000.

5.3.1 Determination of D2S123 allele size distribution in a Scottish cohort

Since short (214bp) and long (228bp) progenitor alleles exhibited significant differences in susceptibility to mutation in the MMR deficient, lymphoblast cell line lbl-1261 (Chapter 4), the size distribution of D2S123 alleles in a Scottish cohort was investigated to address whether large differences in constitutional allele size were observed within a population. Thirty-nine Scottish individuals were genotyped at the D2S123 locus. A total of eight different alleles were identified ranging from 210-230bp. Sizes corresponded to variations in the core number of CA repeats rather than single nucleotide changes. This was consistent with the published CEPH data to which these results were compared (http://www.cephb.fr/cephdb/) (Figure 5.1 and Table 5.1). The 198bp alleles, observed in the CEPH data, were not identified in this cohort. However a previously undocumented allele size of 230bp was observed. The allele sizes in this study (shown) genotyped exactly 1bp longer than the corresponding CEPH alleles. This is likely to be due to differences in primer length.

Table 5.1 Comparison of D2S123 allele frequencies between CEPH database and 39 Scottish individuals. Genotyping data from the small Scottish cohort is in agreement with those from CEPH family analysis. Number refers to that assigned by the CEPH database. A 198bp allele was not observed in the Scottish cohort, however a previously undocumented allele size of 230bp (U) was observed. Allele sizes shown here (a) genotyped exactly 1bp longer than the corresponding CEPH alleles.

Number	Size of Fragment[a] (bp)	Frequency (CEPH data)	Frequency (Scottish data)	Category
1	228	0.304	0.244	Long
2	212	0.339	0.41	Short
3	216	0.107	0.089	Short
4	214	0.125	0.128	Short
5	216	0.054	0.026	Short
6	198	0.018	0.00	Short
7	226	0.018	0.064	Long
8	210	0.036	0.013	Short
U	230	0.00	0.026	Long

D2S123 alleles were noticed to cluster into two distinct size groups, long (~228bp) or short (~214bp) and the frequency of alleles in each of these two groups were almost identical between the Scottish cohort analysed here and the CEPH data (Figure 5.1 and 5.2, and Table 5.1). In addition, it was noted that short alleles predominate over long alleles in both the CEPH and Scottish cohorts (Figure 5.2)

Therefore the distinct long and short constitutional alleles in cell line lbl-1261 (Chapter 4) reflect two distinct allele size classes of D2S123, detected in the population. Inherent differences in these two allele classes other than that of length, may therefore result in differential mutational stability of the type observed in cell line lbl-1261.

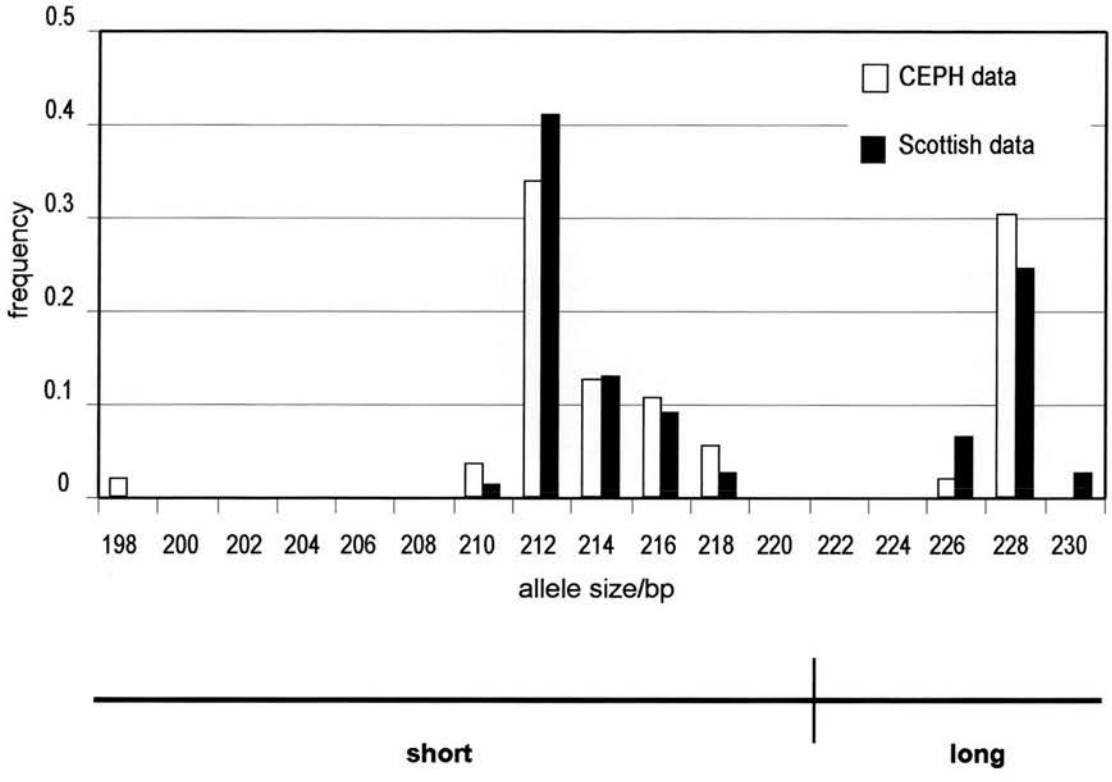


Figure 5.1 D2S123 allele size frequencies in 39 Scottish individuals and the CEPH database. The striking correlation in the frequencies of allele sizes between these two cohorts is clearly illustrated.

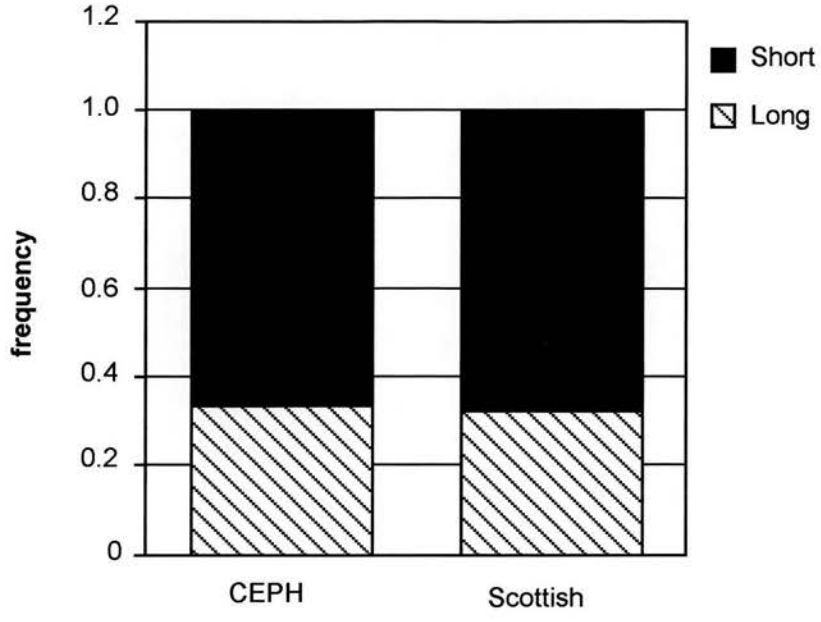
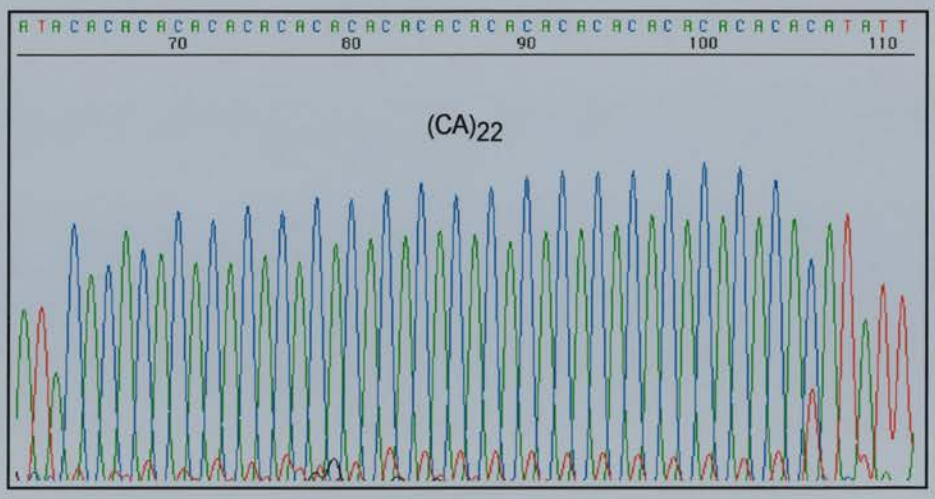


Figure 5.2 Comparison of frequencies of long and short D2S123 alleles in the Scottish cohort and the CEPH database. The frequencies of the two allele classes are almost identical between these two cohorts, with short alleles predominating.

5.3.2 Mutation rate within the D2S123 CA repeat is predicted by sequence content

The sequence for D2S123 logged in GenBank is $(CA)_{13}(TA)(CA)_{15}$ (GenBank accession number; Z16551). Individual alleles from cell line lbl-1261 were sequenced to determine whether sequence differences between long and short alleles might explain the substantial mutation bias. Initially undiluted lbl-1261 DNA amplified at the D2S123 locus was gel separated and individual alleles sequenced independently. This demonstrated that the 228bp allele comprised a $(CA)_{28}$ repeat tract split into two runs of $(CA)_{13}$ and $(CA)_{15}$ by a TA dinucleotide concurring with the sequence logged in GenBank. However, the shorter 214bp allele comprised an uninterrupted $(CA)_{22}$ repeat with no interspersing TA dinucleotide (Figure 5.3). Genotyping and sequencing of 15 D2S123 alleles from template DNA of a cohort of Scottish individuals showed that various length alleles classified as "long" always contained the TA interspersion while, various lengths of "short" alleles were invariably uninterrupted poly CA repeats (Table 5.2). This analysis indicated that alleles clustered around 214bp (short) contain perfect uninterrupted CA repeats. In contrast long alleles consistently contain a TA interspersion within the CA tract (Figure 5.3 and Table 5.2).

A.



B.

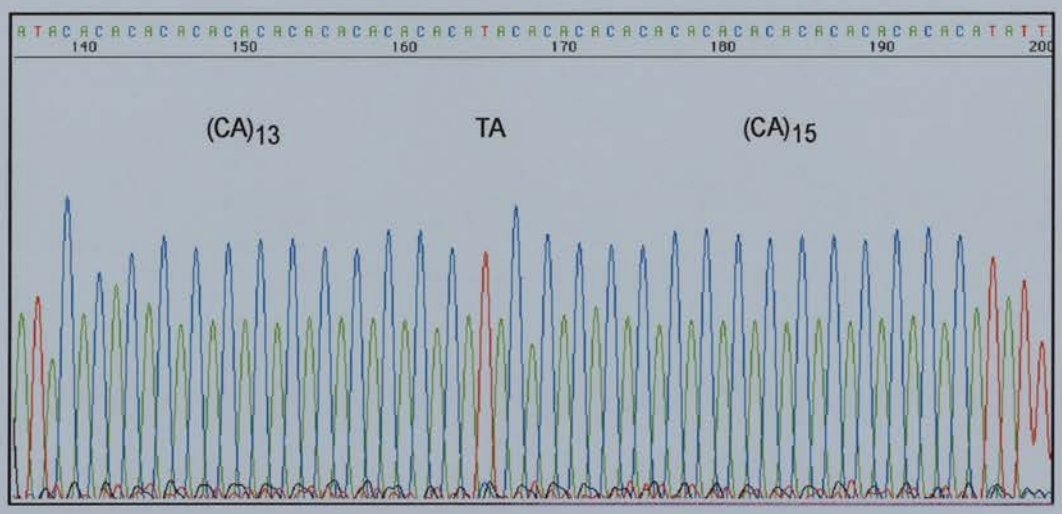


Figure 5.3 Sequence analysis of the progenitor D2S123 alleles cloned from lbl-1261. The short allele (A) consists of an uninterrupted $(CA)_{22}$ tract. The long wild type allele (B) consists of $(CA)_{13}$ repeats followed by a TA dinucleotide and then another run of $(CA)_{15}$ repeats.

Table 5.2 Analysis of D2S123 repeat in 15 alleles from 9 unrelated Scottish individuals. Constitutive allele sizes were determined using an ABI310 and classified as either short (S) or long (L). Sequencing of individual constitutive alleles revealed the presence (interspersed) or absence (perfect) of a TA interspersion within the (CA) repeat tract. NR no result.

Sample	Allele genotypes (short or long)	Repeat status of CA tract
lbl-c5	212 (S)	perfect
	214 (S)	perfect
lbl-1260	212 (S)	perfect
	228 (L)	interspersed
lbl-162	228 (L)	interspersed
	228 (L)	interspersed
lbl-830	214 (S)	perfect
	228 (L)	interspersed
lbl-2135	216 (S)	NR
	228 (S)	interspersed
lbl-757	212 (S)	perfect
	212 (S)	perfect
lbl-c40	212 (S)	perfect
	218 (L)	NR
lbl-55	216 (S)	perfect
	228 (L)	interspersed
lbl-129	212 (S)	perfect
	216 (S)	NR

In order to confirm definitively that mutant alleles detected in lbl-1261 (Chapter 4) arose predominantly from uninterrupted alleles, D2S123 alleles were PCR amplified from 100ng lbl-1261 template DNA and cloned. Individual alleles were then sized and sequenced. Analysis of alleles from 28 clones confirmed that every long allele analysed (wild type or mutant) contained the TA interruption and every short allele (wild type or mutant) contained perfect CA repeats, thereby confirming that the presence of a TA interspersion confers stability on the CA repeats (Table 5.3 and Figure 5.4).

Mutation was due to length variation of the number of core CA repeats in every case, implicating replication slippage as the causative mechanism of instability at this

locus (Table 5.3 and Figure 5.4). Cell line lbl-c5 was heterozygous for two short alleles and PCR cloning of undiluted DNA revealed two short uninterrupted alleles of $(CA)_{21}$ and $(CA)_{22}$ repeats. As expected in this MMR proficient cell line, no mutations were detected in the clones despite both alleles having no interspersion of the repeat tract.

Taken together these data indicate that long alleles (~228bp) invariably contain a TA interspersion in the repeat while short alleles (~214bp) contain uninterrupted $(CA)_n$ repeats that are inherently more unstable. This instability was unmasked by the analysis of MMR defective cell line lbl-1261 in Chapter 4.

Table 5.3 Sizing and sequencing of D2S123 alleles cloned from MMR deficient cell line, lbl-1261. The sequences of the CA repeat region only are indicated. [a]Size of alleles as determined by ABI 310 analysis. The allele sizes of 208-216 nucleotides are classed as short and those of 224-230 nucleotides as long (see text and Chapter 4). All short alleles contain perfect CA repeats in contrast to long alleles that have a TA interspersion. Variation in allele size is wholly accounted for by differences in the number of CA repeats.

Allele Size[a]	Sequence	Number of Clones	Category
208	$(CA)_{19}$	3	Short
210	$(CA)_{20}$	6	Short
212	$(CA)_{21}$	4	Short
214	$(CA)_{22}$	5	Short
216	$(CA)_{24}$	2	Short
222	$(CA)_{12}TA(CA)_{13}$	1	Long
226	$(CA)_{12}TA(CA)_{15}$	1	Long
228	$(CA)_{13}TA(CA)_{15}$	5	Long
230	$(CA)_{14}TA(CA)_{15}$	1	Long

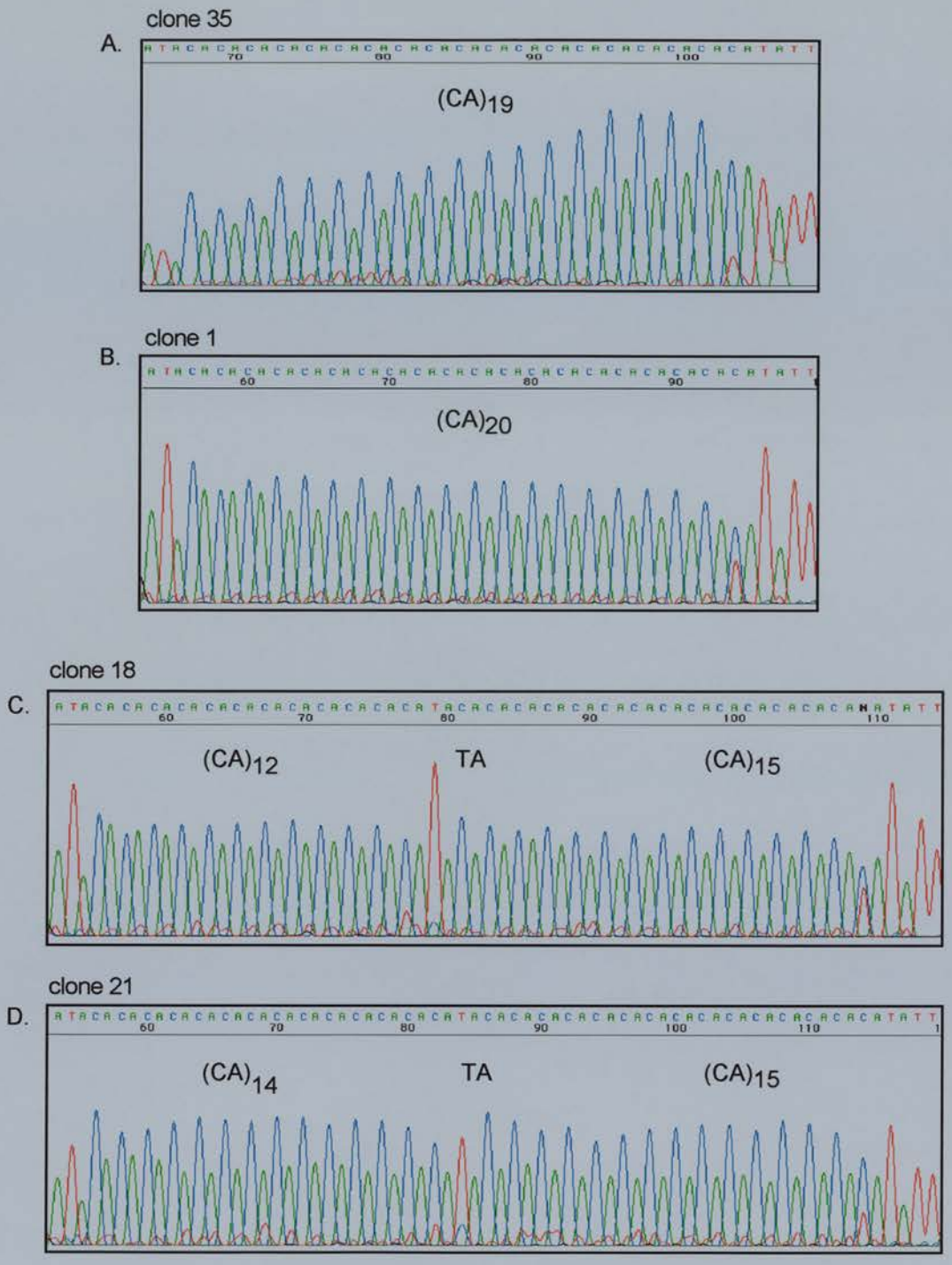


Figure 5.4 Representative sequences of mutant D2S123 alleles cloned from Ibl-1261 DNA. Clones 35 (A) and 1 (B) were genotyped as short with alleles sizes of 208 and 210bp respectively. Both have an uninterrupted CA repeat tract. Clones 18 (C) and 21 (D) were 226 and 230bp in length and thus classified as long. The TA interruption is evident in both.

5.3.3 Allele specific bias of mutations at D2S123 in DNA MMR deficient tumours

To determine whether allele bias might also influence the MSI phenotype in MMR deficient tumours, D2S123 alleles were genotyped in matched normal and tumour DNA samples from 41 colorectal cancer patients with MSI^+ tumours. These have been characterised previously and exhibit MSI at 4 or more markers. In some cases the germline mutation has been identified (Farrington *et al.*, 1998) (unpublished data). Normal tissue was genotyped for D2S123 alleles and the stability of each allele assessed in matched tumour DNA (Figure 5.5). Thirty of the 41 tumours exhibited mutation at one or more D2S123 alleles and overall 37 of the 82 alleles (45%) had mutations (Figure 5.6). There was a significant bias in the frequency of mutations at the short alleles compared with long alleles (Figure 5.6). Of 53 short wild type alleles identified, 32 (60%) displayed instability in the tumour tissue compared with only 5 of 29 (17%) long alleles (χ^2 =12.4, $p<0.001$). Thus, constitutional allele sequence is a determinant of the propensity for instability at a given locus in the presence of defective mismatch repair. These data have important implications for the classification of clinical material with respect to MSI status.

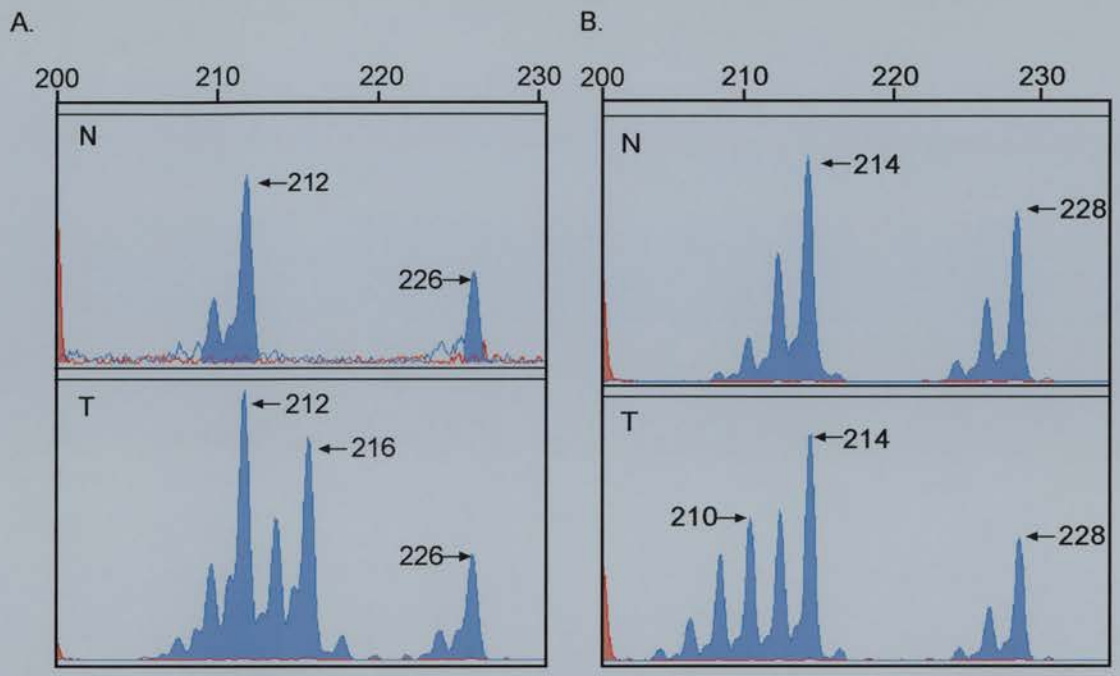


Figure 5.5 Representative ABI 310 profiles of matched normal (N) and tumour (T) DNA from two MSI^+ patients. Patients (A) and (B) both show a shift at the short constitutive allele in the tumour DNA. Both long alleles are stable in the examples shown here.

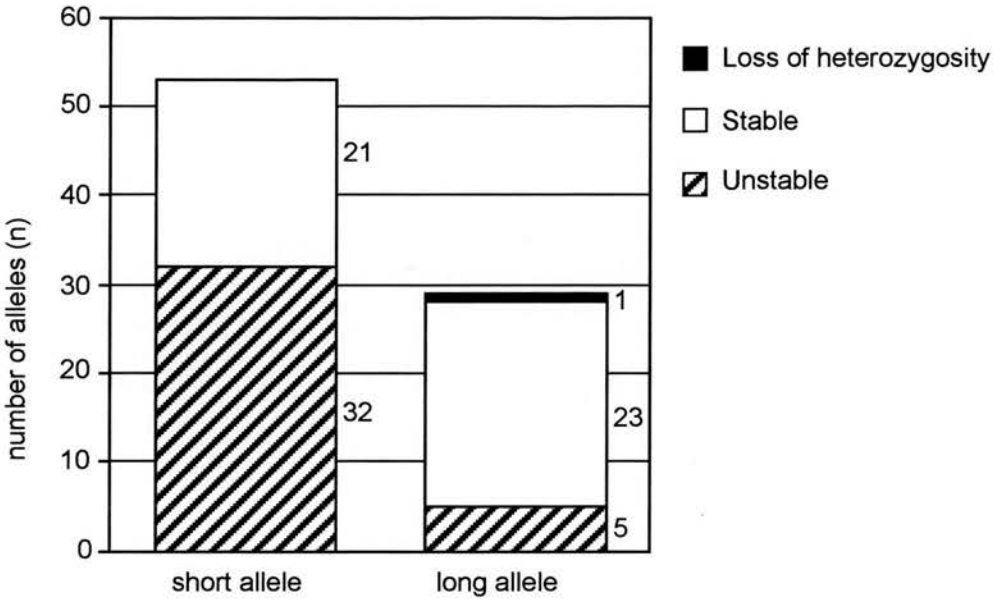


Figure 5.6 Tumour DNA instability at D2S123 alleles in 41 MSI^+ CRC patients. Long and short alleles were compared between matched normal and tumour DNA and assesed for stability. There is a significant excess of mutations at short alleles (χ^2=12.4, p<0.001).

5.4 Discussion

The data presented in this chapter investigate further, the initial observation made in Chapter 4, that intra-allelic bias for mutational events occurs in a MMR deficient cell line derived from normal tissue. The nature of the bias has been determined and demonstrates that intra-allelic sequence variability for an interruption within the repetitive tract, determines allele stability. Furthermore this analysis demonstrates that accumulation of mutations at D2S123 is predicted by host genotype at each allele.

By genotyping and sequencing of individual alleles, it has been demonstrated that defective MMR results in insertion and deletions of CA repeat units, which accounts for the variation in mutant allele sizes observed by SP-PCR in Chapter 4. In addition, alleles without a TA interspersion within the repeat tract are significantly more susceptible to instability as a consequence of defective MMR.

Other studies that have addressed differences between loci in human populations and in CRCs, also show that perfect repeats are more susceptible to instability (Goldstein and Clark, 1995; Petes *et al.*, 1997; Brinkmann *et al.*, 1998). However this study goes further, by demonstrating that variant interruptions can occur between alleles at the same marker loci and that these have a marked effect on individual allele stability. These results indicate that the stabilising effect of a variant repeat is apparent, even in the presence of defective MMR, supporting a previous report (Petes *et al.*, 1997). The data from this chapter and Chapter 4 provide compelling evidence that different alleles at the same marker loci can display genotypic variation, substantially affecting susceptibility to mutation. A factor independent of tumourigenesis, namely allelic heterogeneity for a repeat interspersion has been specifically shown to influence the manifestation of the mutator phenotype.

Several studies have suggested that the mechanism of microsatellite mutation in MSI^+ tumours shares similarity with the evolution of such sequences in the genome (Di Rienzo *et al.*, 1998; Sturzeneker *et al.*, 2000). D2S123 population allele frequencies from the Scottish cohort analysed here and that of the CEPH database highlights higher frequencies of short alleles. The CEPH database,

(www.cephb.fr/cephdb/), documents 6 short D2S123 alleles compared with 2 long ones (Table 5.1, Figure 5.1). This would be expected if short alleles with perfect repeats are more susceptible to mutation, as the results presented here in MMR defective cells suggest. These mutations could then be fixed as new alleles. The absence of many length variants of the long alleles is also consistent with the sequence data here that demonstrate that long alleles invariably contain the stabilising interspersion

It is generally accepted that replication slippage is the major mechanism causing new mutations in microsatellites (Levinson and Gutman, 1987). The results are consistent with models in which a variant interruption causes stabilisation by encouraging the perfect realignment of the two strands, following their dissociation during DNA replication slippage. However the possibility that interruption of the repeat tract may alter some unusual structure associated with the repeats and subsequently reduce the rate of slippage, cannot be wholly discounted (Petes *et al.*, 1997).

The majority of mutations occurring at CA microsatellites have been shown to involve small length changes in which one or two repeats are altered (Henderson and Petes, 1992; Petes *et al.*, 1997; Brinkmann *et al.*, 1998). Mutations of repeat tracts in cancer genes of MSI^+ colorectal tumours characteristically display small frameshift mutations (Rampino *et al.*, 1997; Parsons *et al.*, 1995b). The mutations identified (in Chapter 4) at the D2S123 locus, also involve the gain or loss of one or two repeat units in the majority of cases. In the cell line lbl-1261, short mutant alleles always contained perfect repeats, whereas long mutant alleles consistently possessed a TA interruption. The absence of any short mutants with a repeat interruption, or long mutants without an interruption or of mutants with duplicated TA interspersions, argues against the occurrence of large sequence alterations in excess of a few repeats. This is consistent with PMS2, the MMR gene mutated in lbl-1261, being involved in the repair of small insertion:deletion loops (Kolodner and Marsischky, 1999).

These data show that host genotype at CA repeat sequences, can influence the ability of available marker sets to assign MSI status to any individual tumour. Therefore these findings have important clinical relevance regarding MSI screening

strategies and the effects of individual patient genotype on these analyses. Tumour MSI status is used to determine if HNPCC may be discounted or if analysis of MMR genes is required. Additionally, there have been a number of reports indicating MSI status may be used as a predictor of survival and can be employed as a clinical tool with which to give patient survival estimates and to predict responsiveness to therapeutic treatment (Lothe *et al.*, 1993; Fink *et al.*, 1998; Gryfe *et al.*, 2000). An accurate assessment of MSI status is therefore of critical importance. The other four microsatellite markers in the recommended panel (Boland *et al.*, 1998) are entered in GenBank as uninterrupted $(CA)_n$ or $(A)_n$ repeats. However the CEPH database (www.cephb.fr/cephdb/) indicates that they are polymorphic for different sized alleles. Such analysis by length may mask further underlying individual genotype differences at these markers. Detailed characterisation of these microsatellite markers will be of clinical relevance and of interest in establishing the extent of this intra locus mutational bias. In any case, the phenomenon described here for D2S123 alone is of importance and the influence of patient genotype at this particular marker may be particularly critical in the diagnosis of borderline MSI cases especially when using a small panel of markers. The data from these experiments highlights the need for a well-characterised set of diagnostic markers in which allelic variance and the intrinsic effect on stability is well understood. Suggestions that one or a few microsatellite markers are sufficient to assess MSI status should perhaps be treated with caution, especially when there may be significant implications for both the patient and the family. (Boland *et al.*, 1998; Frazier *et al.*, 1999; Perucho, 1999). The marker D2S123 is logged in GenBank as an interspersed microsatellite. However, it has been shown that this does not reflect the underlying complexity of this locus. Many other markers may share similar intrinsic variances in allelic stability and the phenomenon may be widespread.

Finally, these studies have demonstrated that investigation into the consequences of MMR defects in cells derived from normal tissue can unmask factors at repetitive sequences that can dramatically affect mutation frequency in the absence of MMR.

Chapter 6

Hypermutability at the BAT-40 poly(A/T) Tract in the Human Germline

6.1 Introduction

In Chapter 4, analyses of MMR deficient cell lines lbl-1260 and lbl-1261 revealed that the mutator phenotype exhibits heterogeneity on number of different levels. Heterogeneity in mutation frequency was revealed between cell lines with different MMR gene defects and between different alleles at the same microsatellite locus. In addition, heterogeneity was observed between different microsatellites with a statistically significant elevation in mutation at poly(A/T) microsatellite BAT-40, compared with (CA)n microsatellite, D2S123 (4.3.4). This suggests that BAT-40 is inherently more susceptible to mutation. This could be due to factors such as the extreme length of the repeating unit, as well as the nature of the repeat and these influences have been demonstrated previously to contribute to instability (Weber and Wong, 1993; Sia *et al.*, 1997; Wierdl *et al.*, 1997; Zhang *et al.*, 2001).

BAT-40 is highly sensitive to the effects of defective mismatch repair, since it is susceptible to mutation in >95% MSI^+ tumours and thus used routinely in the analysis of MSI (Parsons *et al.*, 1995b; Dietmaier *et al.*, 1997; Boland *et al.*, 1998). In addition previous studies have revealed that BAT-40 exhibits significant polymorphism within populations (Zhou *et al.*, 1997; Samowitz *et al.*, 1999). Along with the data presented in Chapter 4, these observations suggest that BAT-40 is inherently highly mutable. It has been demonstrated that mechanisms generating mutations in microsatellite unstable (MSI^+) tumours can have relevance to understanding evolution of such sequences in the germline (Sturzeneker *et al.*, 2000). Therefore since inherent instability at this locus is suggested from its high mutation frequency in MMR deficient cells, BAT-40 might also be unstable in the germline.

Investigation of determinants of germline mutation at SSR loci is laborious and frequently requires analysis of many hundreds of meiosis in family studies (Weber and Wong, 1993; Brinkmann *et al.*, 1998). However the development of SP-PCR

techniques to study germline stability at minisatellites, has facilitated investigations of such mutations at other SSRs (Jeffreys *et al.*, 1994; May *et al.*, 1996).

To date many studies of germline mutation at SSR loci are focused on understanding sequence instability of the trinucleotide repeat disorders (Mornet *et al.*, 1996; Kunst *et al.*, 1997; Monckton *et al.*, 1999; Crawford *et al.*, 2000). In addition a number of studies have investigated germline stability of dinucleotide repeat markers, including long CA stretches (Weber and Wong, 1993). However there has been little investigation of the stability of mononucleotide tracts in the germline. This is surprising since poly(A/T) tracts are the most abundant simple repetitive motif in the human genome, largely due to the poly(A/T) tails of scattered retrotransposed sequences such as Alu, Line-1 and processed pseudogenes (Toth *et al.*, 2000; International human genome sequencing consortium, 2001). Coding poly(A/T) sequence tracts have been identified with repeat lengths of up to 27bp and within introns they may reach up to 70bp in repeat length (Toth *et al.*, 2000). Any process influencing the fidelity of replication at coding sequence mononucleotide tracts, will clearly have important functional effects. For example the *TGFBR2* gene contains a poly(A/T)$_{10}$ tract in exon 3 and has been shown to be mutated in up to 90% of MSI^{+} tumours resulting in inactivation of the gene (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b; Lu *et al.*, 1996; Markowitz, 2000).

In this chapter the stability of BAT-40 in the germline was investigated to determine if the extreme inherent instability at this locus suggested by analysis of MMR deficient cells, could also be revealed in non-somatic tissue. The microsatellite BAT-40 is a paradigm mononucleotide marker consisting of 40 adenine repeats located in intron 2 of the 3-beta-hydroxysteroid dehydrogenase (*3-beta-HSD*) gene on chromosome 11. The stability of this paradigm poly(A/T) tract has been determined by study of population allele frequencies, mutation frequency in families and mutation frequency in sperm DNA.

It is of further importance that BAT-40 is used in the analysis of MSI in colorectal tumours. It has already been demonstrated, in Chapter 5, that inherent influences at such loci can direct their propensity for instability in the presence of MMR defects (Bacon *et al.*, 2000). Such studies highlight the need for markers used in these

assessments to be extensively characterised in order for them to be employed with confidence. Assessment of the degree of mutability at a locus such as BAT-40 might have considerable relevance to the generation of mutations at that locus in MMR deficient tumours.

6.2 Methodological Overview

6.2.1 Biological material

Genotyping was carried out on the constitutional DNA of 102 unrelated Scottish individuals as described in 2.3.2, and 35 unrelated CEPH family individuals from the nine families described in 2.3.1.

A Scottish Family, K-435, was used for pedigree analysis and is described in 2.3.5. DNA from the 9 CEPH families, described in 2.3.1 was used for further pedigree analysis to provide a total of 187 germline transmissions for study.

SP-PCR was carried out on matched constitutional and sperm DNA samples from two further unrelated individuals described in 2.1.5.

6.2.2 BAT-40 genotyping

Genotyping of constitutional DNA was carried out in triplicate by amplifying BAT-40 alleles using HEX labelled BAT-40F and BAT-40R primers as described 2.4.1 and the Expand High Fidelity PCR system. These primers amplify a 126bp product containing the standard 40(A)s according to the genomic sequence of the *3-beta-HSD* gene (GenBank; M38180) (Lachance *et al.*, 1990). It was not possible to assess the number of adenine repeats in a given sized allele directly, since repeated attempts at sequencing across the BAT-40 locus were unsuccessful. Therefore repeat length is based on the theoretical predicted amplified sequence both in this and in the majority of other studies that report BAT-40 repeat length (Zhou *et al.*, 1997; Samowitz *et al.*, 1999; Yokozaki, 2000).

PCR products were size analysed using an ABI 310 Genetic Analyser as described in 2.5.1.

6.2.3 Sperm DNA preparation and SP-PCR

DNA was extracted from sperm as described in 2.2.2 and 2.2.3 and from matched constitutional DNA as described in 2.2.1. Limiting dilution experiments, SP-PCR and analysis of products was carried out as described in 2.4.4 and 2.4.5. using BAT-40 primers as above. Avoidance of contamination is paramount when preparing DNA samples and amplifying dilute DNA templates and stringent measures against contamination were taken as described in 2.5.4.

6.2.4 Origin of new alleles

Assessment of allele origin was carried out according to 2.5.4.

6.2.5 Statistical analysis

Statistical comparison of population allele size frequency at the BAT-40 locus was carried out using a Mann Whitney U test on the Minitab (V.13) statistical package. Significance was taken at the 5% level.

For comparison of mutation frequency between matched sperm and blood samples, the frequency of mutant alleles in each sample was expressed as the number of alleles which were mutant in length divided by the total number of alleles detected (normal and mutant). Accordingly the frequency of mutants was not the exact number of cells with alterations but represent relative proportions of alleles. Statistical analyses were then performed using a chi-squared analysis on Minitab (V.13) statistical package, and significance taken at the 5% level.

6.3 Results

The results presented in this chapter have been published in Bacon *et al.*, 2001a.

6.3.1 The BAT-40 poly(A/T) locus is polymorphic

For any inherently unstable locus, rapid generation of new alleles would be expected to result in appreciable levels of polymorphism within a population. Therefore, BAT-40 genotypes were defined in 104 unrelated Scottish individuals and 35 unrelated CEPH family members (Table 6.1). Representative ABI310 profiles are shown in Figure 6.1. PCR products displaying single complex peaks with a near normal distribution were counted as homozygous (Figure 6.1A). Those with extra peaks were regarded as BAT-40 heterozygous (Figure 6.1B-D).

Table 6.1 Levels of heterozygosity at the BAT-40 locus. Frequency of individual BAT-40 genotypes in the Scottish and CEPH cohorts analysed are indicated. Allele sets are given in bps and are grouped according to whether or not the genotype is heterozygous.

	Allele set /bps	Freq. detected in Scottish cohort/% , n=104	Freq. detected In CEPH cohort/%, n=35
Heterozygotes	108/119	0	2.9
	108/123	1.9	0
	109/124	0	2.9
	108/125	1	0
	109/122	1	0
	111/123	0	2.9
	111/124	1	0
	112/124	1	0
	118/119	1	0
	118/121	2.9	0
	118/123	1.9	0
	118/124	1	0
	119/120	1	0
	119/121	1.9	0
	119/122	6.7	5.7
	119/123	1	2.9
	119/124	1.9	0
	119/126	1	0
	119/127	1	2.9
	120/121	1	0
	120/122	5.7	11.4
	120/123	9.6	8.6
	120/124	1.9	5.7
	120/125	1.0	0
	121/122	1.0	0
	121/123	3.8	5.7
	121/124	0	2.9
	122/123	0	2.9
	122/124	1.9	0
	122/125	1.0	0
	123/124	1.0	0
	123/125	1.9	0
	123/128	1	0
	124/125	0	2.9
	124/127	0	2.9
	124/134	1	0
Homozygotes	117/117	1.9	0
	118/118	2.9	0
	119/119	4.8	0
	120/120	3.8	8.6
	121/121	3.8	0
	122/122	4.8	0
	123/123	16.3	22.9
	124/124	3.8	5.7

The allele traces of the BAT-40 mononucleotide marker are complex with "stutter" peaks evident due to DNA polymerase slippage. However, "bona fide" allele sizes are taken to be the predominant peak in each separate peak complex in accordance with previous studies (Samowitz *et al.*, 1999; Yokozaki, 2000). The predominant peak is that with the greatest peak area as indicated by the ABI310 genetic analyser software. Analysis of BAT-40 by SP-PCR of multiple single alleles, in three individuals in Chapter 4 has also validated this method of allele sizing at the BAT-40 locus. The most predominant peaks as genotyped from constitutional DNA are detectable as individual alleles by SP-PCR (Figure 4.2 and 4.6). If the true allele size is not that of the predominant ABI310 peak in SP-PCR products, then it must be that almost every allele amplified from each sample has undergone PCR error to exactly the same degree (Table 4.1 and Figure 4.6). This is an unlikely explanation particularly in view of the allele profiles from MMR proficient cell lines lbl-c5 and lbl-c1 where almost no mutant alleles were detected.

PCR error is evident when amplifying BAT-40, by nature of the stutter bands that are observed (Figure 6.1). However, reproducibility of the prominent peaks in a given individual assures confidence in the sizes given (see methodology). Where the genotype of an individual could not be confirmed by reproducibly detecting the same peaks in the allele trace, these individuals were discarded from analysis (3 cases).

The distribution of BAT-40 heterozygous genotypes indicates that amplification and detection of two BAT-40 alleles of different sizes, is due to the difference in genotype and not technical artefact (Table 6.1).

Allele frequency and distribution for both the Scottish and the CEPH cohorts is shown in Figure 6.2. Of the 139 samples analysed, a total of 83 demonstrated heterozygosity at the BAT-40 locus (59.7%). Levels of heterozygosity were similar between the cohorts, 58.7% (61/104) in the Scottish population and 62.9% (22/35) in the CEPH cohort (Table 6.1). As expected for a highly polymorphic marker, the overall distribution of allele sizes was not significantly different between these two populations ($p=0.056$) (Figure 6.2). Allelic size variation was from –15 to +11 as compared to the most frequent allele. Taking into account variation in the cohorts studied, these data are in line with a previous study reporting polymorphism at this

mononucleotide marker for a large number of different alleles (Zhou *et al.*, 1997; Samowitz *et al.*, 1999) and are suggestive of the frequent generation of new alleles at this locus. However, the BAT-40 heterozygosity reported here differs in both frequency and size distribution to that observed in a Japanese study despite the use of similar methodology (Yokozaki, 2000)

The most frequent allele in both cohorts in this study (123bp) corresponds to a 37-adenine tract as calculated from the genomic sequence (GenBank; M38180).

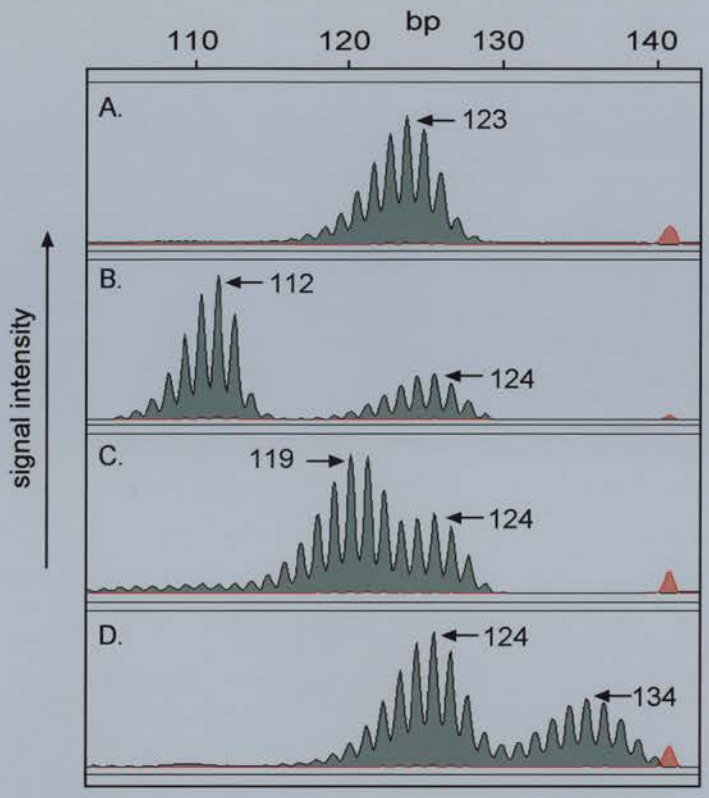


Figure 6.1 Representative ABI 310 traces of BAT-40 poly(A/T) PCR products. The polymorphic nature of this microsatellite marker in samples from a Scottish cohort is clearly demonstrated. Blood DNA (A) shows a single complex of peaks the maximum being 123bp. Samples B-D show heterozygosity at BAT-40 as illustrated by seperate peak complexes.

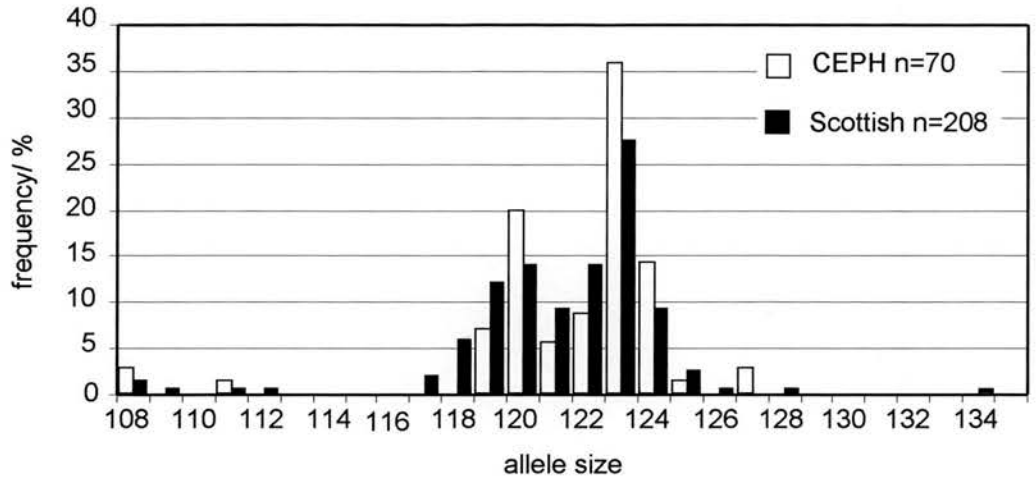


Figure 6.2 Comparison of BAT-40 allele frequencies between Scottish and CEPH populations. Allele sizes are given in bp. The most frequent allele size (123bp) corresponds to a BAT-40 allele containing 37 adenine repeats as estimated from the predicted PCR product size (126bp) of the standard BAT-40 allele with 40 adenine repeats (Genbank; M38180). There is no statistical difference in the distribution of alleles within the two cohorts (p=0.056).

6.3.2 Germline hypermutability at BAT-40 in pedigree analysis

Since BAT-40 displays high levels of polymorphism in populations and has been previously demonstrated to be extremely susceptible to instability in MMR deficient tumours (Liu *et al.*, 1995a; Dietmaier *et al.*, 1997), it was reasoned that BAT-40 may be inherently unstable and that germline mutations might be detectable in family studies. A Scottish family (K-435) was used for pedigree analysis, as Dr. Farrington in the laboratory had previously identified proband MD-473 as heterozygous at BAT-40, with two distinct sized alleles at this locus (112/124) (Figure 6.3). Analysing the meiotic stability of BAT-40 alleles that are easily distinguished by size allows for the most accurate assessment of individual allele stability at a complex locus such as BAT-40. DNA from 20 available individuals from K-435 was genotyped at the BAT-40 locus (Figure 6.3). There were 11 germline transmissions available for study. Within the family there was evidence of a germline mutation in the allele transmission from MD-1303 to MD-449. MD-1303, is heterozygous for BAT-40 with allele set 120/124 but her daughter (MD-449) is homozygous with two 112 BAT-40 alleles (Figure 6.3 and 6.4). DNA was unavailable from the father of MD-449 who is highly likely to have carried a 112 allele, inferred from sibling and progeny genotypes. Therefore, the mutation is implicated as being maternal in origin, displaying loss of repeats at the BAT-40 locus.

The 112/112 homozygous genotype in MD-449 is unlikely to have arisen by dropout of the larger 120 or 124bp allele during PCR, because four family members including MD-439, the sister of MD-449, had 112/124 genotypes that were easily detected under the PCR conditions used (Figure 6.4). This indicates that the technique reliably detects larger alleles. Furthermore, the 120bp allele in MD-1303 was faithfully amplified (Figure 6.4). Genotyping for all members of this family were confirmed in triplicate and previous genotyping at microsatellite markers confirmed that MD-1301 was indeed the mother of the twins MD-449 and MD-439 (Dr. S. Farrington, pers. comm.).

This observation of a germline mutation in only 11 transmissions is striking, since only one mutation event in 3000-5000 transmissions has been reported for CA repeats (Weber and Wong, 1993). These data provide founding evidence for the

notion that BAT-40 is inherently prone to mutation and is unstable in the germline. This initial observation led to further investigation into the possibility that BAT-40 is highly unstable in the germline and that mutant alleles might be transmitted in subsequent generations, to manifest as polymorphism within the population.

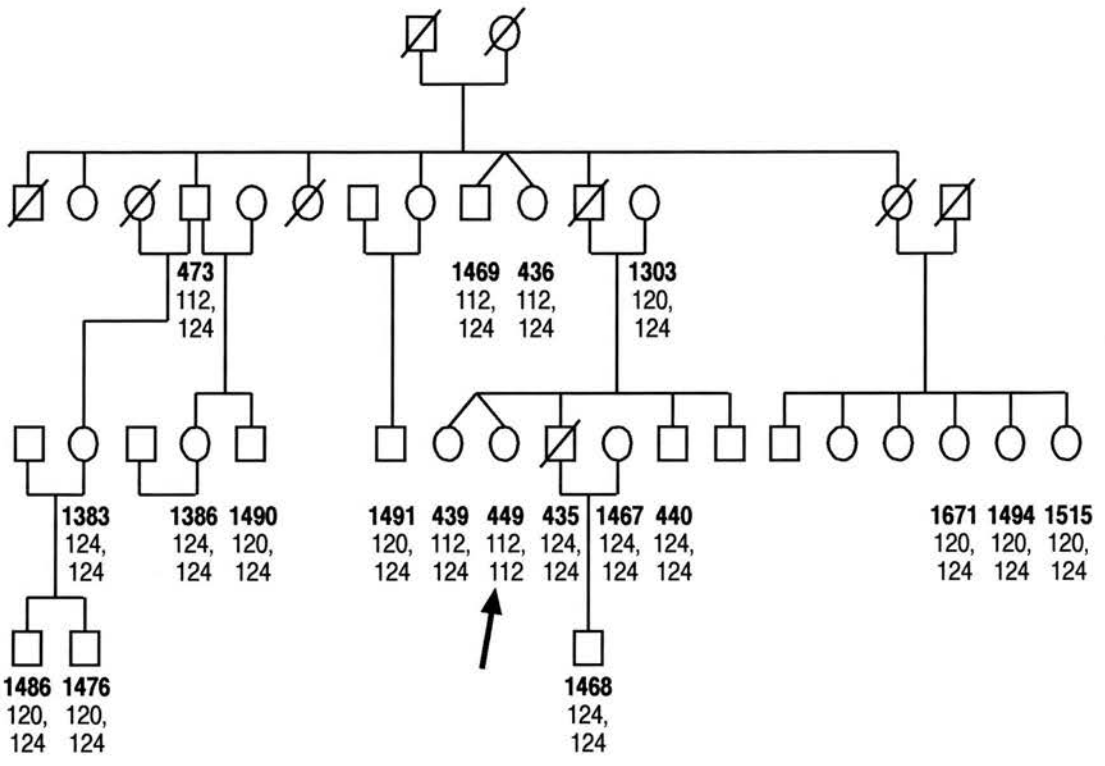


Figure 6.3 BAT-40 genotypes of blood DNA from HNPCC pedigree K-435. A single incidence of transmissible germline hypermutability is highlighted. Mother MD-1303 has allele set 120/124 where as daughter MD-449 (arrow) is homozygous for 112.

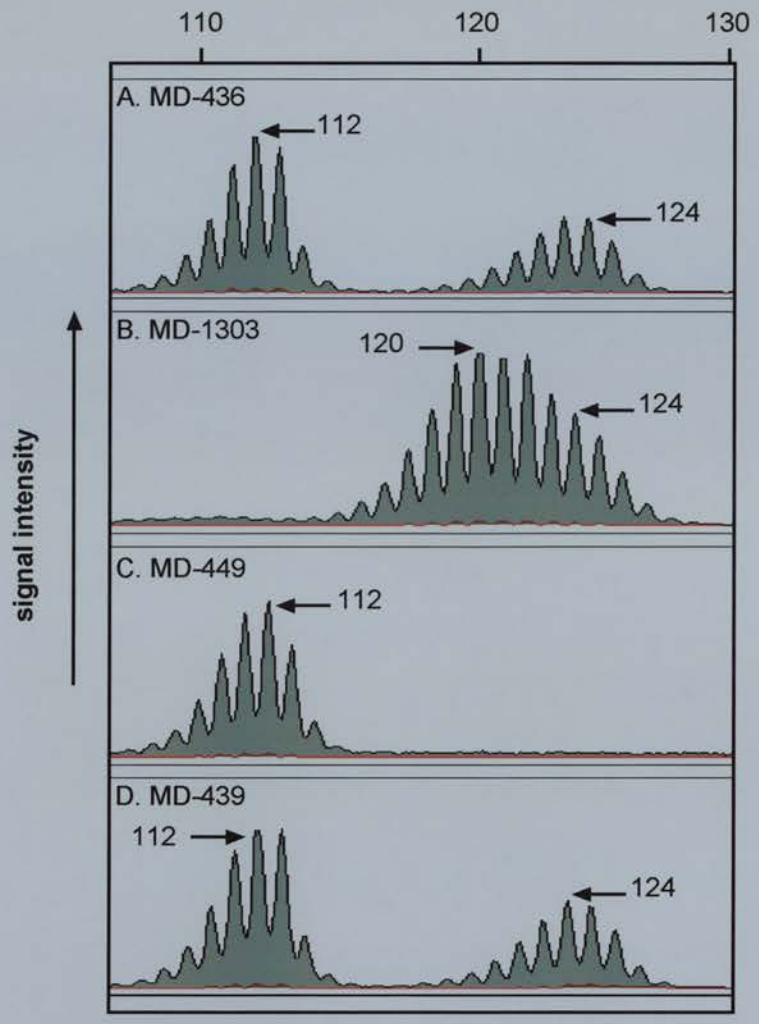


Figure 6.4 Germline hypermutability in pedigree K-435. ABI 310 profiles of BAT-40 genotypes are shown. (B) MD-1303 is the mother of MD-449 and alleles of 120 and 124bp are detectable. (C) There is no indication of the presence of either of MD-1303s alleles in MD-449, who is homozygous for allele 112.(A) Although DNA was not available from the father it is likely that he carried at least one 112 allele as inferred from siblings such as his sister, MD-436. (D) The profile of MD-439, who is the twin sister of MD-449, also demonstrates that there is not a problem in detecting the 124 allele in the presence of the 112 allele.

6.3.3 Germline hypermutability at BAT-40 in CEPH family analysis

To further explore germline instability at BAT-40, BAT-40 alleles were analysed in a CEPH family panel. Nine CEPH families were genotyped at the BAT-40 locus, totalling 176 germline transmissions and 12 putative mutations (6.8%) were identified (Table 6.2 and Figure 6.5). It should be noted that 11 BAT-40 alleles were counted as non-mutated by inference, due to the failure of two DNA samples to amplify.

Table 6.2 Putative BAT-40 mutations detected in germline transmissions of CEPH families. Transmissions in which a germline mutation was indicated are described and the genotype of the child along with both parents is presented. The inferred origin is highlighted with an asterisk. In one case (a) the sample failed to amplify and genotype was inferred from the relatives. The CEPH pedigree number, individual identification number and family relationship are given in standard CEPH nomenclature.

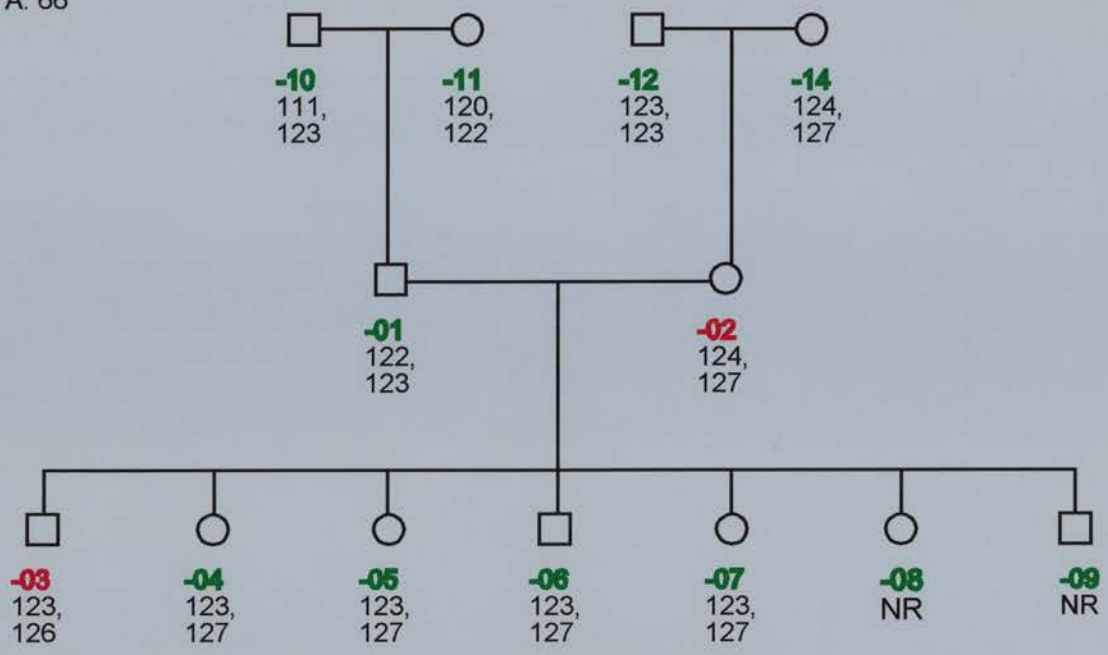
Family	Genotype (father)	Genotype (mother)	Genotype (child)
66	122/123 (f, -01)	124/127 (m, -02)*	123/126 (c/m, -03)
	123/123 (fm, -12)*	124/127, (mm, -14)	124/127 (m, -02)
1331	123/123 (f, -01)*	119/123 (m, -02)	119/124 (c/m, -17)
1341	120/123 (f, -01)*	123/125 (m, -02)[a]	119/125 (c/f, -05)
	120/123 (f, -01)*	123/125 (m, -02)[a]	119/125 (c/f, -08)
1346	123/123 (f, -01)*	122/127 (m, -02)	124/127 (c/f, -08)
	123/123 (f, -01)*	122/127 (m, -02)	124/127 (c/f, -09)
1362	120/120 (fm, -15)	121/123 (mm, -16)*	120/120 (m, -02)
	121/123 (f, -01)*	120/120 (m, -02)	120/120 (c/f, -04)
1377	120/124 (ff, -10)	120/122 (mf, -11)*	120/121 (f, -01)
	120/121 (f, -01)*	119/122 (m, -02)	122/123 (c/m, -08)
13293	120/123 (f, -01)*	108/109 (M, -02)	109/124 (c/m, -09)

However, in all cases, heterozygous parental alleles differed by only a few base pairs and the mutations indicated involved small (1bp) changes (Table 6.2 and Figure 6.5). Figure 6.6 illustrates an example of one of the BAT-40 germline mutations detected in family 1346.

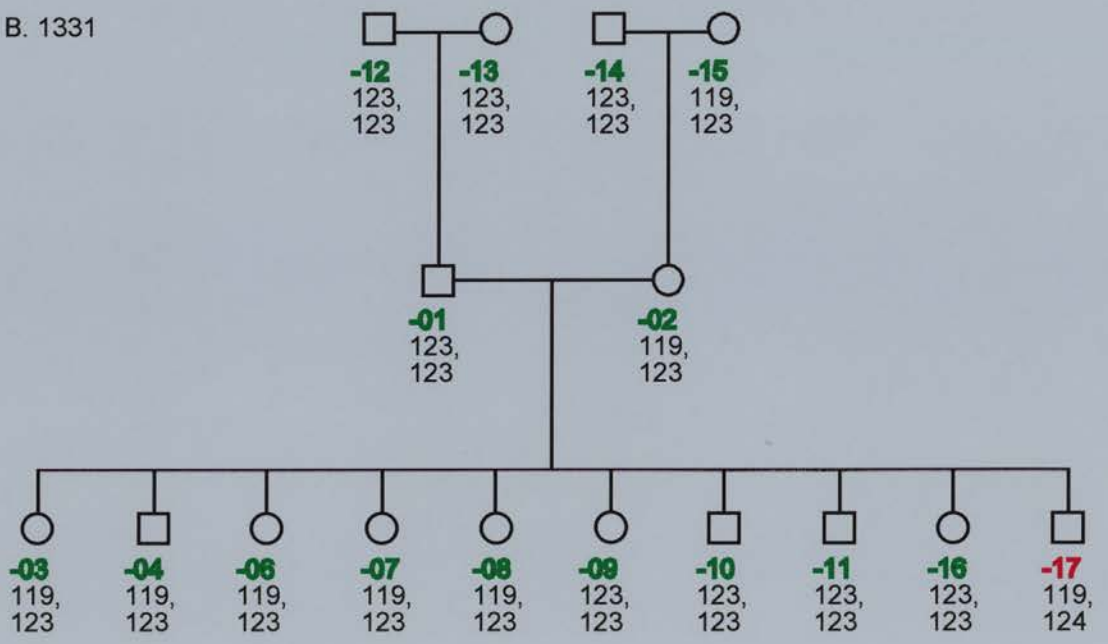
Of the 88 maternal transmissions analysed, 3 were mutant at BAT-40 (3.4%) and of the 88 paternal transmissions analysed 9 were mutant (10.2%). This difference was not statistically significant ($\chi 2=3.22$, $p=0.073$). Insertions and deletions appeared to occur equally.

The mutations identified in CEPH families provided further support for the initial observation that BAT-40 is hypermutable in the germline. The CEPH data adds further weight to the identification of the germline mutation in family K-435 and also to the evidence from the high levels of heterozygosity at BAT-40 in the Caucasian population study, demonstrating high levels of heterozygosity at BAT-40. However, although the CEPH mutations were reproducible, the small changes observed in the complex BAT-40 profile (Figure 6.6), led to a further, rigorous method being devised to analyse susceptibility of this locus to mutation in the germline.

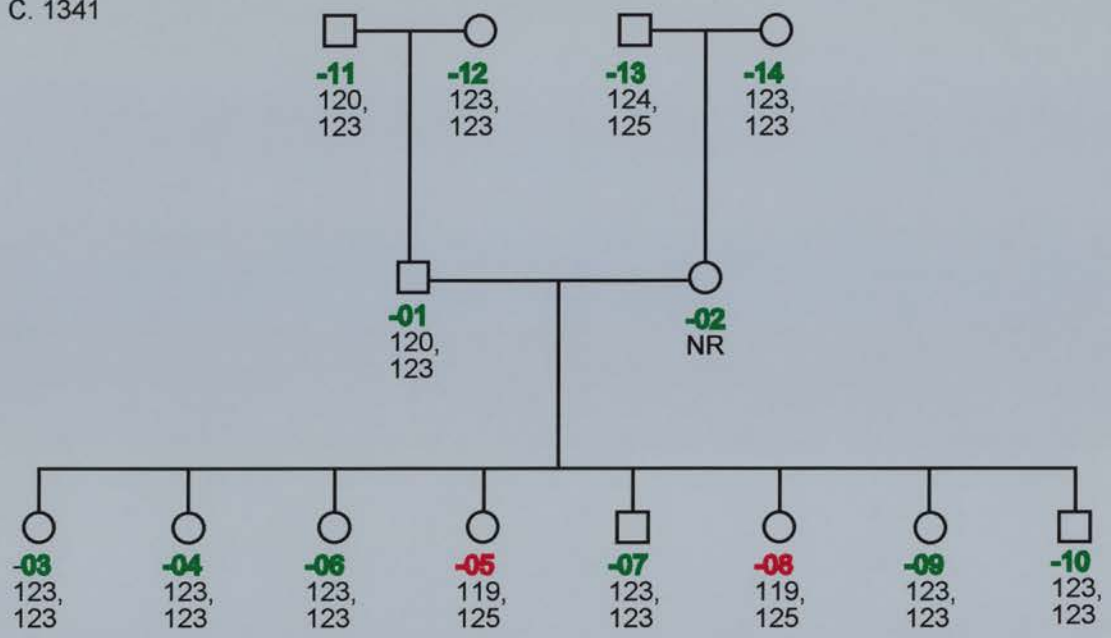
A. 66


-10
111,
123
-11
120,
122
-12
123,
123
-14
124,
127
-01
122,
123
-02
124,
127
-03
123,
126
-04
123,
127
-05
123,
127
-06
123,
127
-07
123,
127
-08
NR
-09
NR

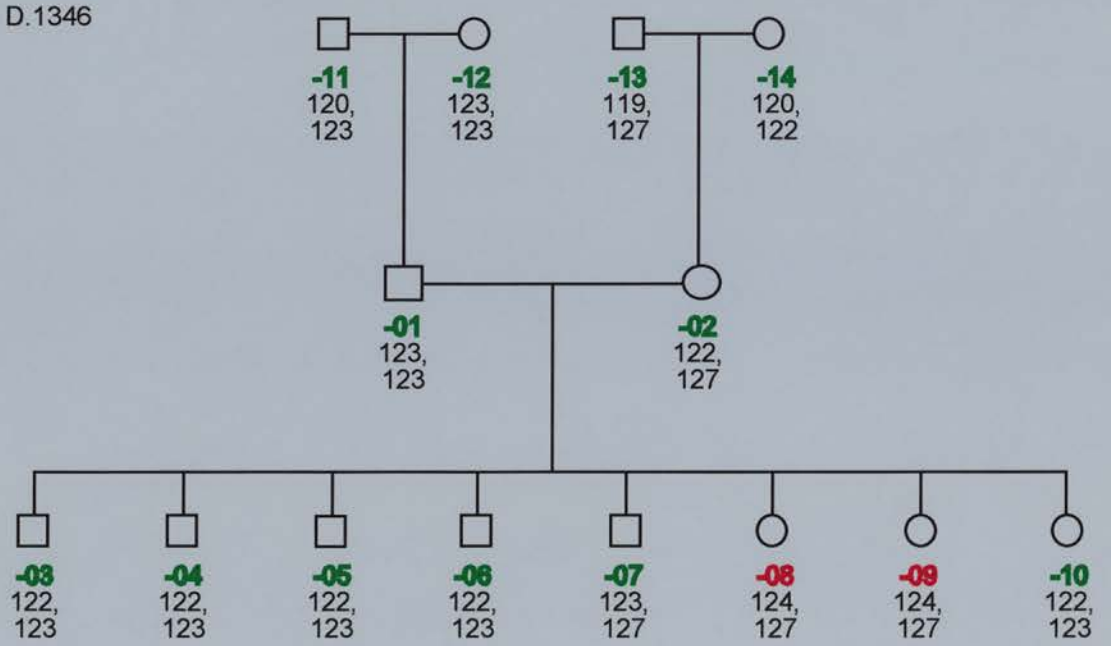
B. 1331


-12
123,
123
-13
123,
123
-14
123,
123
-15
119,
123
-01
123,
123
-02
119,
123
-03
119,
123
-04
119,
123
-06
119,
123
-07
119,
123
-08
119,
123
-09
123,
123
-10
123,
123
-11
123,
123
-16
123,
123
-17
119,
124

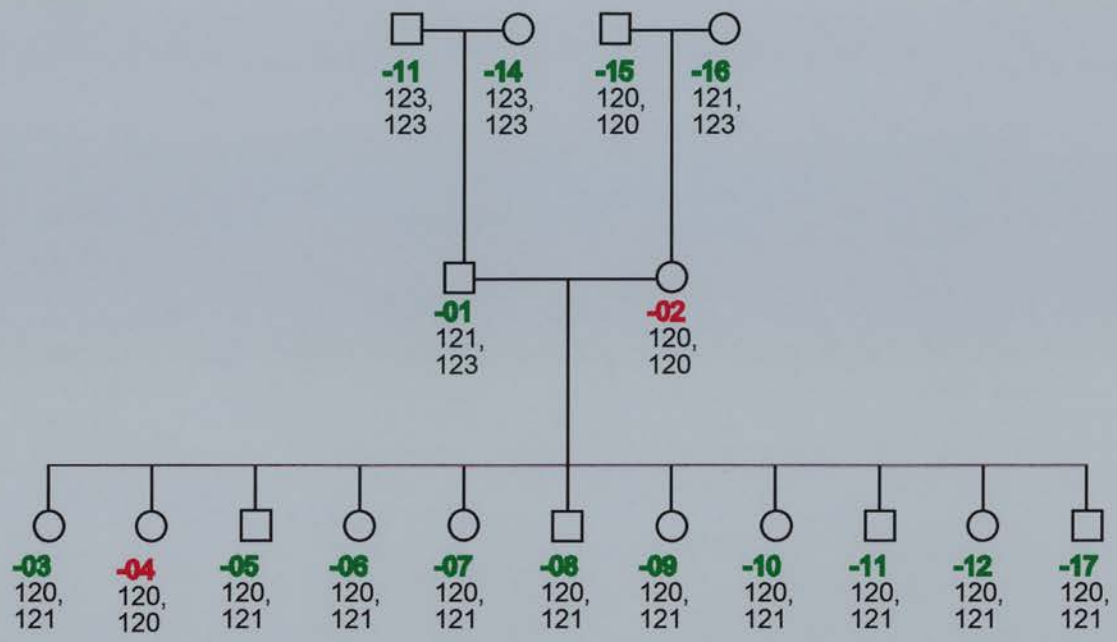
C. 1341


-11
120,
123
-12
123,
123
-13
124,
125
-14
123,
123
-01
120,
123
-02
NR
-03
123,
123
-04
123,
123
-06
123,
123
-05
119,
125
-07
123,
123
-08
119,
125
-09
123,
123
-10
123,
123

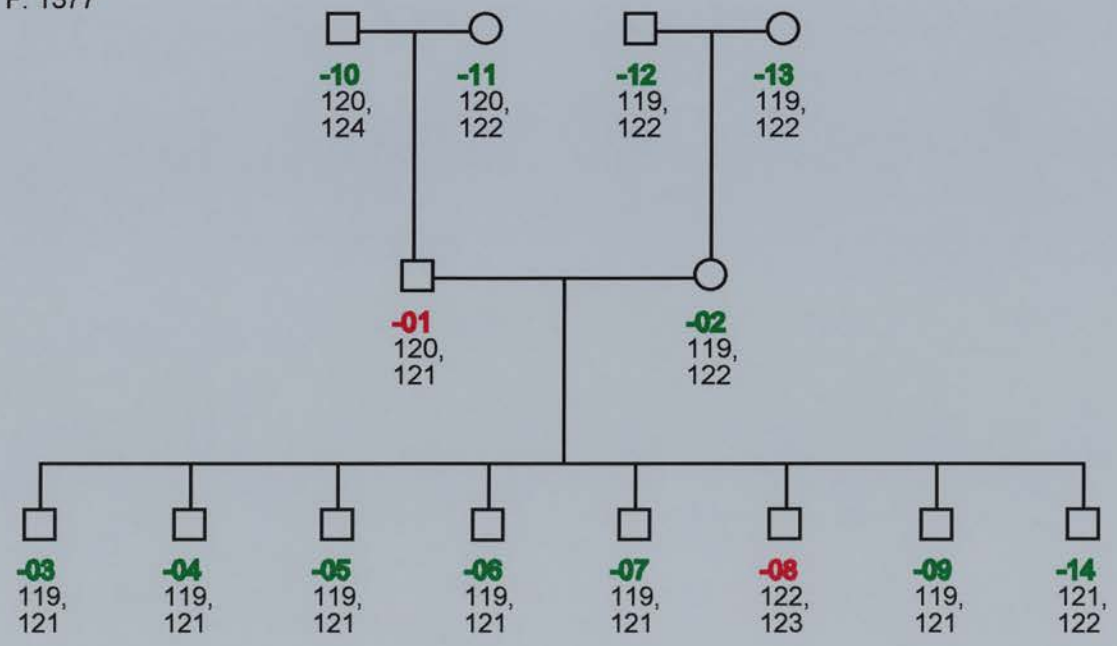
D.1346


-11
120,
123
-12
123,
123
-13
119,
127
-14
120,
122
-01
123,
123
-02
122,
127
-03
122,
123
-04
122,
123
-05
122,
123
-06
122,
123
-07
123,
127
-08
124,
127
-09
124,
127
-10
122,
123

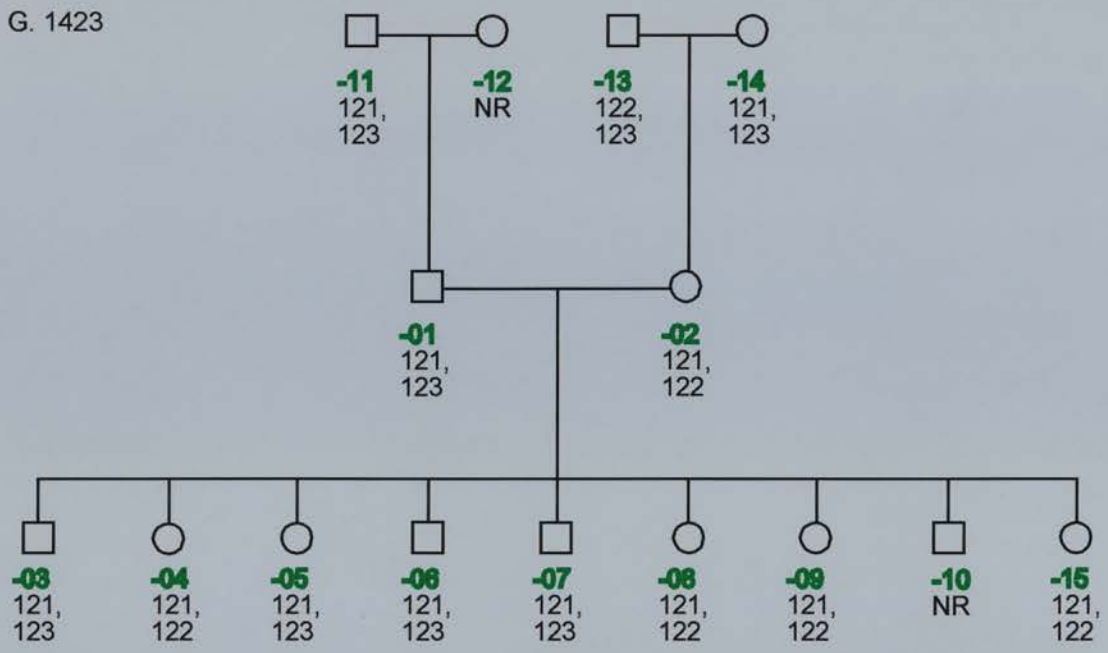
E. 1362


-11
123,
123
-14
123,
123
-15
120,
120
-16
121,
123
-01
121,
123
-02
120,
120
-03
120,
121
-04
120,
120
-05
120,
121
-06
120,
121
-07
120,
121
-08
120,
121
-09
120,
121
-10
120,
121
-11
120,
121
-12
120,
121
-17
120,
121

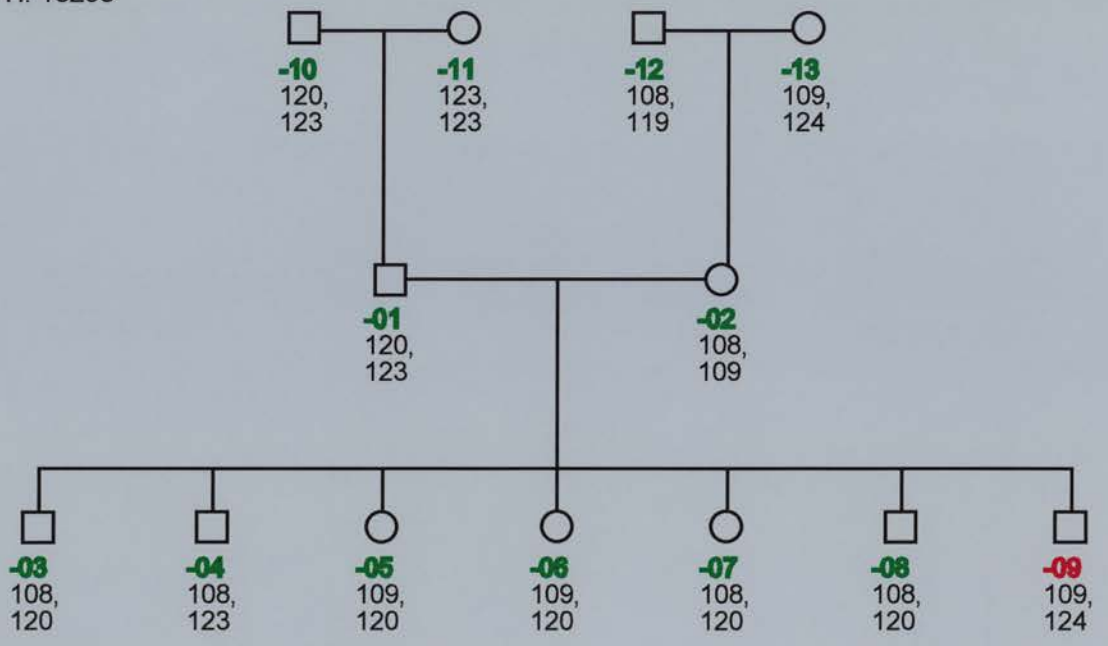
F. 1377


-10
120,
124
-11
120,
122
-12
119,
122
-13
119,
122
-01
120,
121
-02
119,
122
-03
119,
121
-04
119,
121
-05
119,
121
-06
119,
121
-07
119,
121
-08
122,
123
-09
119,
121
-14
121,
122

G. 1423



H. 13293



I. 13294

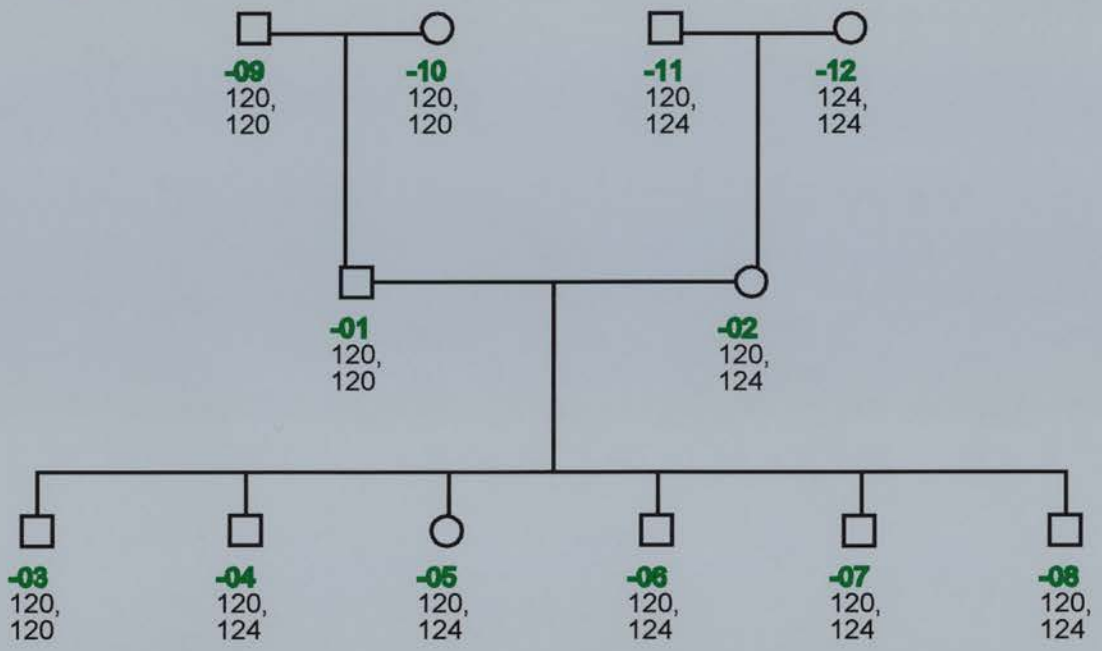


Figure 6.5 Genotyping of the BAT-40 locus in 9 CEPH families. Individuals from CEPH families; 66 (A), 1331 (B), 1341 (C), 1346 (D), 1362 (E), 1377 (F), 1423 (G), 13293 (H) and 13294 (I) were genotyped at the BAT-40 locus. Both allele sizes are indicated for each individual in bp. Identifying CEPH numbers are shown in green or where a putative mutation was identified in red. In a few cases DNA failed to amplify giving no result (NR).

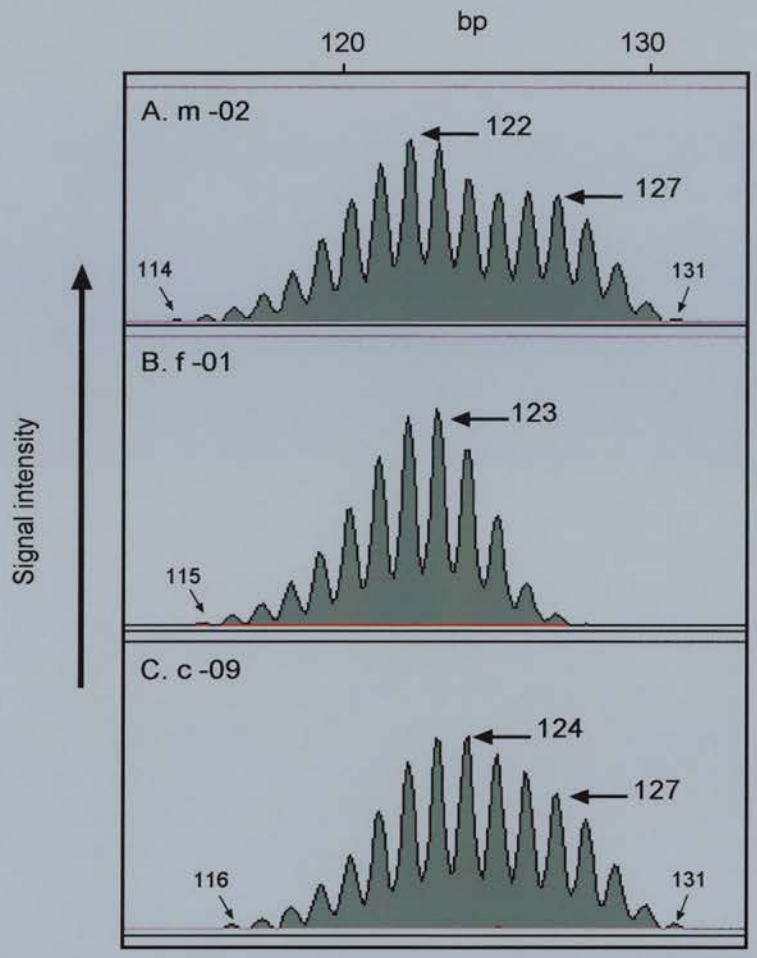


Figure 6.6 Representative example of a putative BAT-40 germline mutation in CEPH family 1346. (A) While the mother -02 has a BAT-40 genotype of 122/127, (B) the father -01 appears homozygous for a 123bp allele. (C) The mothers 127bp allele is detected in child -09 but the most prominent peak in the first complex is at 124bp. This would indicate a 1bp mutation at BAT-40 had occured in the germline of the father. Sizing of the extreme stutter peaks also indicates a 1bp mutation in the fathers 123bp allele in c-09 and confirms the presence of the 127 allele derived from m -02.

6.3.4 SP-PCR analysis of sperm DNA demonstrates that BAT-40 is inherently hypermutable in the germline

SP-PCR analysis of germline DNA has important advantages over family studies for analysing germline stability at complex loci (Jeffreys *et al.*, 1994; Kunst *et al.*, 1997). This method overcomes the practical constraints encountered during pedigree analyses, which suffer limitations from the small number of mutants that can be identified. In contrast, many hundreds of gametes can be analysed from a single semen sample and consequently, greater variations in size that can be easily detected, should be available for identification. In addition the dilution of the DNA sample aids unambiguous identification of both smaller and larger size changes at a hypermutatble locus (Jeffreys *et al.*, 1994).

Mutation frequency as detected in sperm DNA has been shown to reflect estimations from studies in pedigrees (Jeffreys *et al.*, 1994). Comparisons of sperm DNA and constitutional DNA templates has shown that SP-PCR reliably discriminates alleles in both and that there are no demonstrable differences in technical artefact between the two sample templates (Jeffreys *et al.*, 1994). In addition SP-PCR has been demonstrated to reliably detect differences in intra-allelic mutation frequency in sperm DNA (Mornet *et al.*, 1996; Kunst *et al.*, 1997; Jeffreys *et al.*, 1994; Crawford *et al.*, 2000).

For SP-PCR of sperm DNA, study subjects were selected on being constitutionally heterozygous at the BAT-40 locus with individual wild type alleles easily distinguished by size. MD-c1 had allele sizes 120/124 and MD-949 had alleles of size 121/124. Correct identification of constitutional allele sizes was confirmed in the SP-PCR analysis where individual alleles of the same predominant allele size were detected (Figure 6.7). Approximately 100 SP-PCR products were analysed per sample. BAT-40 allele sizes typed from constitutional and sperm DNA templates by SP-PCR are shown in Figure 6.8. Mutant alleles were detected in sperm DNA by comparison to constitutional genotype (Figure 6.7). The frequency of mutant alleles detected in each sample is summarised in Table 6.3.

Table 6.3 Summary of mutant alleles detected by SP-PCR in matched sperm and blood DNA from samples MD-949 and MD-c1.

Sample	Total no. alleles	Mutants (frequency)	
MD-c1 Sperm	164	44 (0.27)	
MD-c1 Blood	99	5 (0.05)	(χ^2 =19.32 p <0.001)
MD-949 Sperm	91	20 (0.22)	
MD-949 Blood	99	4 (0.04)	(χ^2 =13.82, p < 0.001)

In both MD-c1 and MD-949 matched samples, there was a significantly higher number of mutant alleles detected in sperm DNA compared to that of matched blood leukocyte DNA samples (χ^2 =19.32, p<0.001; χ^2 =13.82, p<0.001 for MD-c1 and MD-949 respectively). A total of 9/198 (4.5%) alleles in the leukocyte DNA were mutant compared to a total of 64/255 (25.1%) mutant alleles in the sperm templates, indicating an almost 6 fold increase in mutation accumulation in the germline. The proportion of mutant alleles in blood and sperm was equivalent in MD-c1 and MD-949 samples suggesting that the constitutional heterozygous DNA mismatch repair gene mutation of MD-949, does not influence mutation rate in the male gamete.

Although both addition and deletion mutations were identified, a bias was observed towards repeat losses over gains in the sperm, (χ^2 =11.0, p<0.001; χ^2 =10.3, p<0.001 for MD-c1 and MD-949 sperm samples). While possibly an artefact of the technique, both insertion and deletion mutations were detected in the cell lines lbl-1260 and lbl-1261. These cell lines were evaluated by an identical method to that used here suggesting the technique can detect both larger and smaller mutants.

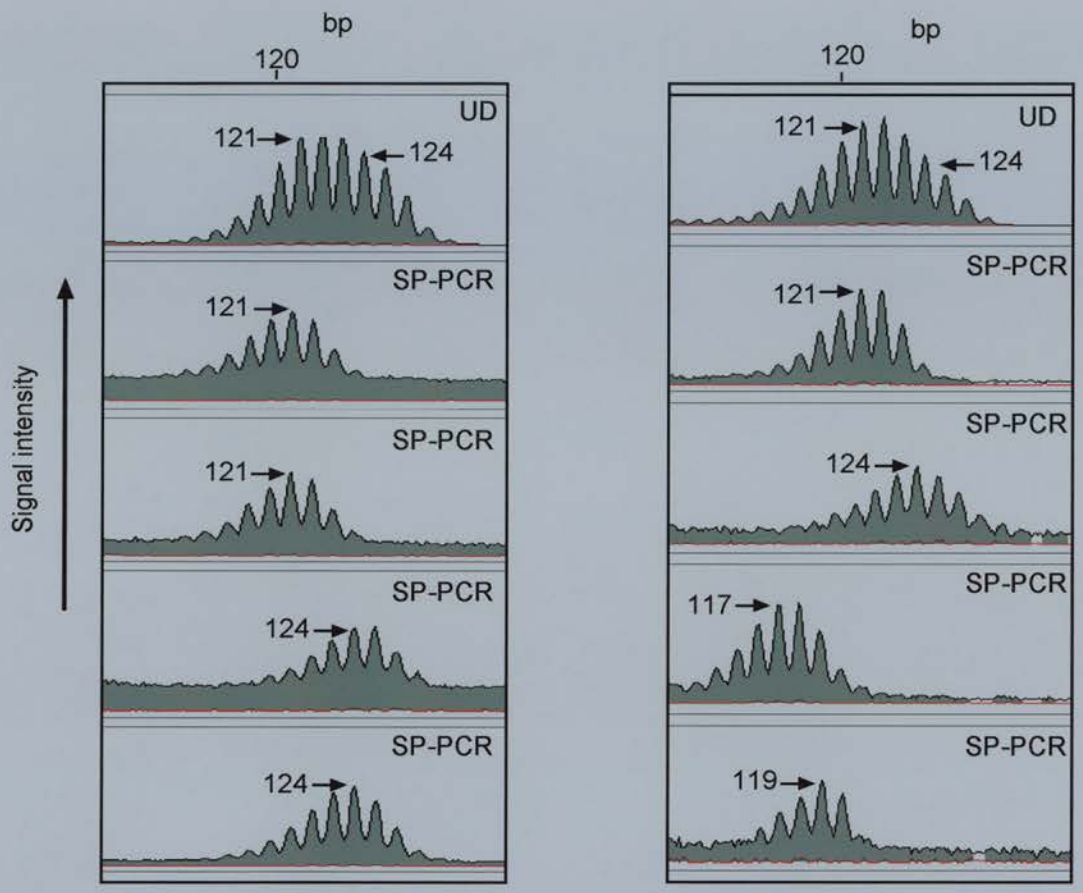


Figure 6.7 Representative ABI310 traces of BAT-40 alleles detected by SP-PCR in matched blood (A) and sperm (B) DNA. (A) Almost all BAT-40 SP-PCR products amplified from blood DNA revealed individual alleles with predominant peaks of the same size as those in the undiluted (UD) DNA. For MD-949 these were 121 and 124bp. (B) The majority of BAT-40 SP-PCR products amplified from sperm DNA were also of wild type allele size as shown. However, a significant number of mutant alleles were detected. Mutants of 117 and 119bp are illustrated.

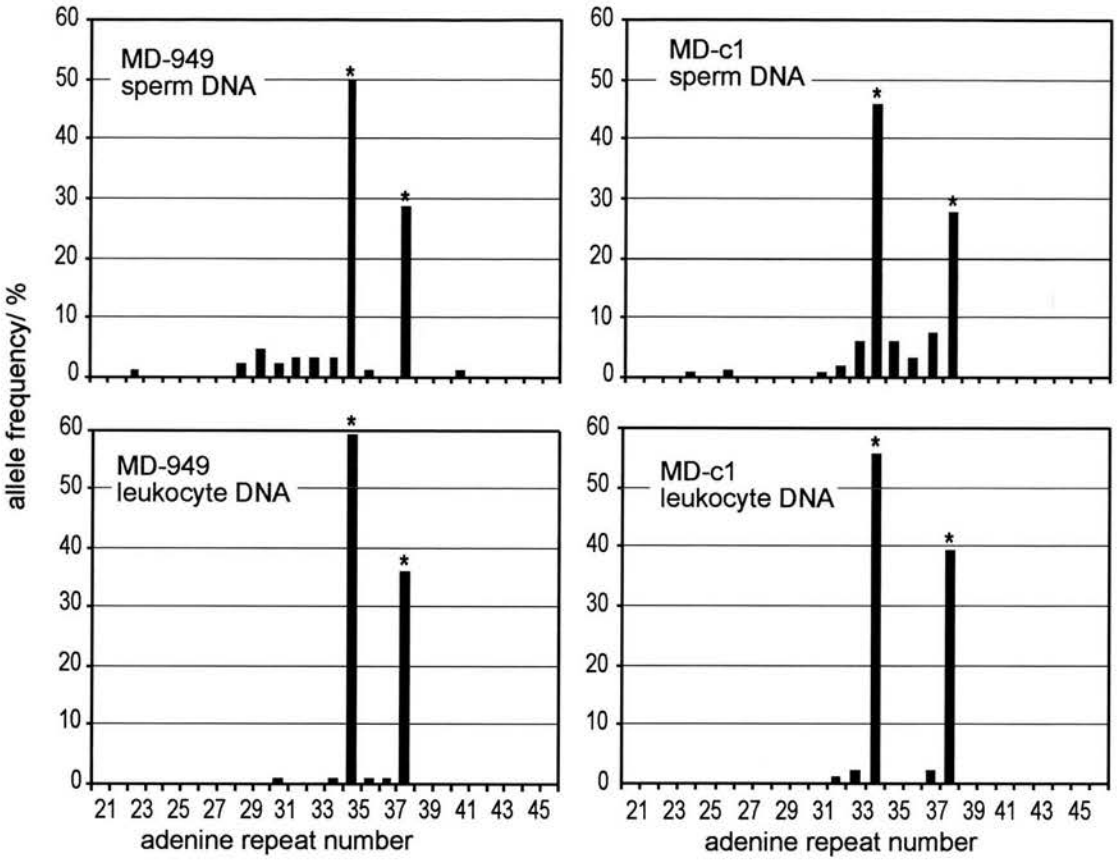


Figure 6.8 BAT-40 allele sizes in matched constitutional and sperm DNA detected by SP-PCR. The predominant allele sizes for each individual as detected from analysis of undiluted DNA are indicated by an asterisk . MD-949 is a CRC patient and has germline mutation in the human *MLH1* gene. MD-c1 is a normal healthy control individual.

6.4 Discussion

This chapter has addressed the hypothesis that BAT-40 is inherently unstable by defining the germline stability of this paradigm poly(A/T) repeat locus.

By analysis of two cohorts it is shown that that BAT-40 is a highly polymorphic locus with an observed level of heterozygosity of 59.7%. This is similar to a previous analysis of a CEPH cohort, in which a level of heterozygosity of 72% was reported ($\chi 2=3.22$, $p=0.073$) (Zhou *et al.*, 1997). However, the level of BAT-40 heterozygosity detected here and by Zhou *et al* is considerably higher than that observed in a Japanese cohort (14.6%) (Yokozaki, 2000). Although this might in part be explained by variation in allele heterozygosity between populations, the Japanese study cohort was from hospital derived samples and may not be representative of the true Japanese population frequencies (Yokozaki, 2000). The confirmation here that BAT-40 is a highly polymorphic locus suggests that generation of new alleles by slippage and mutation at this locus might be quite common.

Further to the data presented in Chapter 4, high levels of instability at BAT-40 are also well-documented in MMR deficient tumours (Parsons *et al.*, 1995b; Dietmaier *et al.*, 1997) suggesting that this marker might be particularly unstable. Identification of a germline mutation at BAT-40 in a Scottish pedigree suggests this locus is also highly unstable in the germline. Failure to amplify larger alleles as an explanation for apparent germline mutations such as that in MD-449, is unlikely. Larger alleles were reliably detected in the presence of the 112 allele in other family members in pedigree K-435. The initial observation in a single family is supported by pedigree analysis of a further nine CEPH families, albeit with less dramatic examples, biased in part by the nature of the parental genotypes. Hence, individuals were chosen with easily distinguishable allele sizes for analysis of sperm DNA. Analysis of matched sperm and blood DNA at BAT-40 by a SP-PCR technique demonstrated a statistically significant increase in the proportion of mutant alleles in sperm compared to somatic DNA. This argues strongly that the mutations detected by SP-PCR of sperm DNA are indeed authentic. In addition rigorous controls were employed to ensure against contamination (see methodology). The SP-PCR approach used in this study has been shown previously to detect mutant alleles with equal

fidelity in sperm and constitutional DNA templates as demonstrated by direct comparisons between mutation rates detected by SP-PCR of sperm compared to family studies (Jeffreys *et al.*, 1994). Previous studies using this technique have consistently validated the SP-PCR approach (Jeffreys *et al.*, 1994; Crawford *et al.*, 2000).

The presence of an inactivating *MLH1* mutation in the germline of one of the individuals did not appear to further influence the level of instability at BAT-40 in the sperm DNA. Since the mutation in MD-949 is heterozygous, this indicates that constitutive loss of MMR is necessary to reveal instability.

Intriguingly, shorter mutant alleles predominated in the sperm DNA despite the fact that the SP-PCR technique reliably detected both large and short constitutional alleles.

The results provide compelling evidence that BAT-40 is inherently unstable in the germline. Hypermutability at the BAT-40 locus provides an explanation for the wide spectrum of allelic variants present in the Scottish and CEPH populations studied, since transmission of new germline variants can become established within the population. The evidence that BAT-40 represents a poly(A/T) tract within the genomic structure of a gene and exhibits instability in the germline, might be of importance in understanding mechanisms generating mutations at other such polymorphic repeat loci. Indeed this phenomenon may be common to many poly(A/T) tracts and further study of such sequences is merited to elucidate whether this is a widespread phenomenon. Poly(A/T) tracts are ubiquitous at the 3'UTR of all coding genes where the stability in length of the poly(A) tail is of known functional importance to the stability of the mRNA species (Bernstein and Ross, 1989). Of further relevance, these repetitive tracts are common in intronic sequence (Toth *et al.*, 2000), and shortening of intronic mononucleotides has been shown to have functional consequences. Aberrant splice variants of the *ATM* gene that result in mutations responsible for ataxia-telangiectasia, are generated as a consequence of shortened intronic mononucleotide tracts (Ejima *et al.*, 2000). A similar phenomenon occurs in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene (Chu *et al.*, 1993). The shortened poly(T)$_5$ variant in intron 8 of the *CFTR* gene causes

congenital bilateral absence of the vas deferens when associated with a cystic fibrosis mutation on the other allele (Chillon *et al.*, 1995). Mutation of poly(A/T) tracts within exonic sequences, has also been shown to contribute to carcinogenesis. This is exemplified by mutation of the *TGFBR2* gene in MSI^+ CRCs (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b; Lu *et al.,* 1996). Hence it seems reasonable to speculate that the mechanism of inherent instability elucidated here, might also have relevance to a number of genes containing such repeats. Similar investigation at other unstable and/or polymorphic loci would reveal the extent of this phenomenon.

BAT-40 is used routinely as a marker in determining tumour genomic stability in relation to defective DNA mismatch repair due to its extreme sensitivity to mutation in the absence of MMR activity (Lothe *et al.*, 1993; Liu *et al.*, 1995a; Rodriguez-Bigas *et al.*, 1997; Boland *et al.*, 1998; Gryfe *et al.*, 2000). Microsatellite markers such as BAT-40 that display extreme germline hypermutability should be used with caution, in view of the likelihood of mitotic instability. Very unstable markers may be too sensitive to provide the specificity to MMR defects that is clearly required in such screening strategies. The evidence reported here supports a growing number of studies, which highlight the importance of understanding inherent characteristics influencing marker stability when they are used in clinical programs (Boland *et al.*, 1998; Thibodeau *et al.*, 1998; Frazier *et al.*, 1999; Samowitz *et al.*, 1999; Bacon *et al.*, 2000).

Chapter 7

Susceptibility of *TGBFB2* and *BAX* Coding Repeat Sequences to Mutation in Non-Cancer Derived cells with MMR Defects

7.1 Introduction

In addition to mutations accumulating at non-coding microsatellite sequences, MMR defects also and more crucially, result in an increased mutation frequency within the repetitive tracts of coding sequence in MSI^+ tumours (Ionov *et al.*, 1993; Strand *et al.*, 1993; Thibodeau *et al.*, 1993; Markowitz *et al.*, 1995; Rampino *et al.*, 1997). A mutator phenotype combined with selective pressure is believed to allow the acquisition of mutations in key genes leading to progression of neoplasia from cellular clones to invasive cancers (Vogelstein *et al.*, 1988; Fearon and Vogelstein, 1990; Loeb, 1994; Kinzler and Vogelstein, 1996; Loeb, 2001). However, the relative importance of selection for growth advantage and genomic instability in tumourigenesis is not clear (Tomlinson, 1999; Sieber *et al.*, 2000; Loeb, 2001). There is little conclusive evidence in favour of either somatic natural selection being the primary force behind tumour growth, or genomic instability being necessary to explain the number of mutations found in cancers (Huang *et al.*, 1996; Sieber *et al.*, 2000; Loeb, 2001). A number of studies have suggested that MSI is an early event in the progression of a normal cell towards a malignant phenotype (Aaltonen, 1994; Shibata *et al.*, 1994; Pedroni *et al.*, 2001). However, the significance and contribution of this event to the accumulation of frequently observed mutations in MSI^+ CRCs is unclear.

The repetitive tracts of the *TGFBR2* and *BAX* genes are mutated frequently in MMR deficient tumours (Parsons *et al.*, 1995b; Markowitz *et al.*, 1995; Lu et al., 1996; Ouyang *et al.*, 1998; Yamamoto *et al.*, 1998; Abdel-Rahman *et al.*, 1999; Markowitz, 2000) and the occurrence of these mutations are believed to be significant events in tumourigenesis (Alexandrow and Moses, 1995; Parsons *et al.*, 1995b; Rampino *et al.*, 1997; Grady *et al.*, 1999). However there are notable

differences in the frequency with which these two genes are mutated in MSI^+ CRCs. While the poly$(A)_{10}$ repeat of *TGFBR2* is mutated in around 90% of MSI^+ CRCs, mutations of the *BAX* poly$(G)_8$ repeat is consistently detected in around 50% of such cases (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b; Rampino *et al.*, 1997; Ouyang *et al.*, 1998; Markowitz, 2000). In addition, the timing of these mutations during tumour progression also appears to be different. Mutations in *TGFBR2* are thought to be an early event, whereas *BAX* mutations seem to occur later (Akiyama *et al.*, 1997a; Grady *et al.*, 1998; Yagi *et al.*, 1998; Abdel-Rahman *et al.*, 1999).

It was hypothesised that differences in mutational frequencies at the *TGFBR2* and *BAX* genes may reflect differences in the relative contributions of inherent mutation and the effects of selection at these loci. Therefore, in this chapter the intrinsic instability of these genes was investigated using cell lines lbl-1260 and lbl-1261 described previously, since MMR is completely defective but the cells are not subject to the potentially extreme mutational bias due to molecular changes and selection pressures that characterise the malignant phenotype. Repeat regions were analysed specifically within *TGFBR2* and *BAX* for mutations and this study provides considerable insight into the inherent stability of two regions mutated frequently in MSI^+ tumours.

7.2 Methodological Overview

7.2.1 Restriction digest assay to detect 1bp deletions in the poly(A)$_{10}$ tract of *TGFBR2*

A restriction digest assay was adapted from that published by Mironov *et al.*, (1999), to provide an initial analysis to detect 1bp mutations in the poly(A)$_{10}$ tract of *TGFBR2* and this is illustrated in Figure 7.1. The assay was carried out as described in 2.5.5 on DNA from lbl-1261 and lbl-1260. DNA from wild type control cell lines, lbl-c5, lbl-c1 and mutant control cell lines, HCT116 and LoVo were also analysed. These assays were carried out in triplicate. In addition, 16 MSI$^+$ tumour DNA samples and 81 constitutional DNA samples from patients with both MSI$^+$ and MSS CRC were analysed in validating the use of this assay and these are described in 2.3.3 and 2.4.4. These assays were carried out in duplicate.

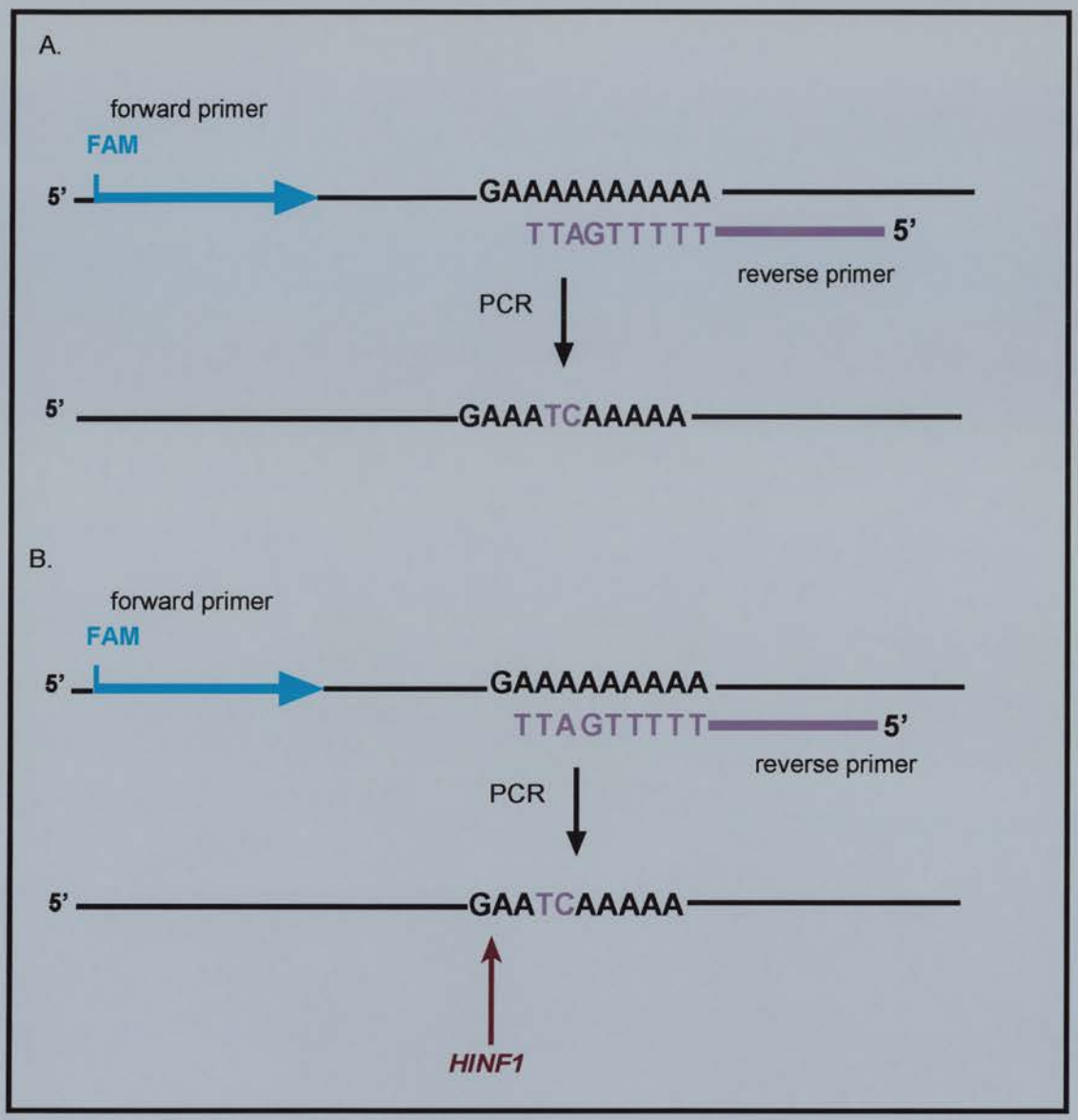


Figure 7.1 Restriction digest assay for the detection of 1bp deletions in the poly(A)$_{10}$ tract of the *TGFBR2* gene. Adapted from Mironov *et al.*, 1999. (A) In the presence of the wild type poly(A)$_{10}$ sequence, no *HINF1* restriction site is introduced during the PCR step. (B) In the presence of a 1bp deletion the reverse primer introduces a *HINF1* site during amplification. For deletion mutants of 2bp or greater amplification would not occur due to increased mis-priming. Restriction digest products are subsequently visualised using an ABI310 genetic analyser. A fluorescent peak at 141bp is detected if the template DNA was wild type (A) and a peak at 118bp will be detected if the template DNA contained a 1bp deletion (B).

7.2.2 Cloning and sequencing

A 266bp region of the *TGFBR2* exon 3 (nt599 –30 to nt789 +46) encompassing the poly(A)$_{10}$ tract present in the *TGFBR2* coding region, (codons 125-128) (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b), was PCR cloned from lbl-1261 and MMR proficient control lbl-c5 using primers TGFBR2F and TGFBR2R as described in 2.4.1, 2.6.1 and 2.6.2. A 245bp non-repetitive region of *TGFBR2* exon 4 was also cloned from lbl-1261 DNA using TGFBex4F and TGFBex4R primers (2.4.1, 2.6.1 and 2.6.2).

Sequencing was performed as described in 2.7.1-2.7.4. All clones were sequenced in duplicate and mutants were re-confirmed by re-isolating and sequencing clone DNA from bacterial stocks. The frequency of false positives and false negatives was either zero or very low for each cell line. Any clone containing a putative mutation that could not be confirmed in duplicate was discarded from the analysis, similarly for any wild type sequence that could not be confirmed in at least duplicate. Sequence data was analysed as described in 2.7.5.

The number of alleles with poly(A)$_{10}$ mutations from each cell line was established and the difference in mutation frequency between cell lines evaluated for significance by a chi squared test as described in 2.10.1. Mutation frequency for individual loci analysed in different cell lines was expressed as follows;

$$\text{Mutation Frequency (mutations/kb)} = \frac{\text{Total number of mutations detected}}{\text{Total sequence analysed/kb}}$$

To test the significance of differences in mutation frequency between different loci or at the same loci in different cell lines, a Chi Squared test or a Fishers exact test was employed as described in 2.10.1. When using this method it is assumed that mutations occur independently at any given base pair within the given locus.

7.2.3 SP-PCR and DHPLC analysis of *BAX* poly(G)$_8$ tract

SP-PCR was carried out on DNA from lbl-1261 and control cell line lbl-c5 using unlabelled BAXF and BAXR primers as described in 2.4.1. These primers amplify a 94bp region encompassing the poly(G)$_8$ repeat at codons 38 to 41 of the *BAX* gene (Rampino *et al.*, 1997). ABI310 analysis was not sensitive enough to convincingly pick up a 1bp deletion in a known mutant due to the small size of this fragment. Therefore 3µl of SP-PCR product was subjected to a second round of PCR, allowing visualisation of products on an agarose gel as described in 2.4.6. These were subsequently analysed by DHPLC using a Transgenomic Wave™ machine as described in 2.5.2 and 2.5.3. A known mutant control (LoVo) was also analysed to ensure mutation detection. Any SP-PCR product resulting in a mutant wave profile was sequenced to confirm the presence of the mutation indicated. In each case, a SP-PCR product with a wild type profile was also sequenced as a control to confirm these were indeed wild type sequences.

7.3 Results

The results presented in this chapter have been published in Bacon *et al.*, 2001b.

7.3.1 Adaptation of a *TGFBR2* restriction digest based assay to detect 1bp deletions in the poly(A)$_{10}$ tract

To address whether the mutator phenotype results in instability within genes known to be involved in colorectal carcinogenesis, the poly(A)$_{10}$ tract of the *TGFBR2* gene was analysed for mutations in lbl-1260 and lbl-1261. Initial analyses demonstrated that the *TGFBR2* repetitive tract was not amenable for SP-PCR analysis, as had been previously employed to detect low level mutation at microsatellite loci. This was due to technical difficulties that were highlighted by the failure to detect mutations in a known mutant cell line, LoVo, by DHPLC or size analysis using an ABI310 genetic analyser. Therefore a sensitive restriction digest assay was adapted to detect 1bp deletions occurring at low frequency within the poly(A)$_{10}$ tract of *TGFBR2* (Figure 7.1) (Mironov *et al.*, 1999). Deletions of 1bp are commonly observed at this repeat tract in MSI$^+$ tumours and thus it was reasonable to use an assay that would specifically detect these lesions (Parsons *et al.*, 1995b Markowitz *et al.*, 1995).

This assay reliably detected the presence of mutant *TGFBR2* alleles in two positive control cell lines (LoVo and HCT116) (Figure 7.2). LoVo has been described previously to have A_9 and A_8 alleles at *TGFBR2* (Carethers and Pham, 2000). This assay revealed only a mutant peak after *HINF1* digestion of *TGFBR2* PCR products amplified from LoVo DNA (Figure 7.2B). Since the *TGFBR2* assay primers amplify alleles of A_9 or longer, (Figure 7.1) only the A_9 mutant allele in LoVo would be amplified and subsequently digested. MMR deficient cancer cell line HCT116 has been previously demonstrated to be homozygous for a 1bp deletion at the poly(A)$_{10}$ repeat of *TGFBR2* (Carethers and Pham, 2000). As expected, a single mutant peak corresponding to the homozygous deletion was detected by this assay

(Figure 7.2C). Finally, a peak corresponding to wild type template was detected by the assay, in wild type control cell lines lbl-c1 and lbl-c5 (Figure 7.2A).

To further validate the use of this assay a panel of 16 MSI^+ CRC DNAs, and a panel of 81 blood DNAs taken from patients with both MSI^+ and MSS colorectal tumours were analysed (Tables 7.1 and 7.2 and Figure 7.2). Mutations were detected in 15/16 (93.8%) of the DNA samples from the MSI^+ tumours (Table 7.1). This frequency is in line with that observed in previous studies using alternative analyses (Markowitz *et al.*, 1995; Markowitz, 2000). No mutations were detected in the constitutional DNAs (Table 7.2), even those from patients (e.g. MD-1699), who had colorectal tumours with identified *TGFBR2* mutations. These preliminary studies indicate the assay adapted for this study can reliably detect known 1bp deletions and that the frequency of mutations observed in a panel of un-analysed MSI^+ tumours is in line with previously published frequencies.

Table 7.1 Summary of *TGFBR2* poly$(A)_{10}$ restriction digest analysis on DNA from MSI^+ CRCs. Tumour identifying numbers are shown. The assay indicated that tumours were either wild type (wt) at this region or that all or nearly all of the cells from the tumour material were homozygous for a 1bp deletion (1bp del.). (See also Figure 7.2)

Tumour ID	Tumour *TGFBR2* poly$(A)_{10}$ status
1478	1bp del
t3	1bp del
t5	1bp del
830	1bp del
t8	1bp del
t9	1bp del
1699	1bp del
1521	1bp del
t12	1bp del
763	1bp del
528	1bp del
t19	1bp del
t20	1bp del
1370	1bp del
t26	1bp del
1339	wt

Table 7.2. Summary of *TGFBR2* poly(A)$_{10}$ restriction digest analysis of constitutional DNA from patients with CRC. One healthy control individual was also analysed. Laboratory identifying MD number is given. The MSI status of patients tumours was determined previously by Dr. S. Farrington and is given as microsatellite stable (MSS), microsatellite high (MSI-H), microsatellite low (MSI-L) or where analysis had not been carried out, not done (ND).

Patient MD-	Constitutional DNA, *TGFBR2* poly(A)$_{10}$ status	Tumour MSI status	Patient MD-	Constitutional DNA, *TGFBR2* poly(A)$_{10}$ status	Tumour MSI status
106	wt	MSS	1385	wt	MSS
177	wt	MSS	1407	wt	ND
329	wt	MSI-H	1442	wt	MSS
432	wt	MSS	1452	wt	MSI-L
528	wt	MSI-H	1459	wt	MSS
559	wt	MSI-H	1460	wt	MSS
611	wt	MSS	1462	wt	ND
615	wt	MSI-H	1464	wt	ND
621	wt	MSS	1465	wt	ND
622	wt	MSS	1509	wt	ND
633	wt	MSI-H	1513	wt	MSI-L
649	wt	MSS	1523	wt	MSS
660	wt	MSI-H	1583	wt	MSS
682	wt	MSS	1596	wt	ND
696	wt	MSI-H	1597	wt	MSS
708	wt	MSS	1603	wt	MSI-H
737	wt	MSS	1622	wt	MSI-H
757	wt	MSS	1628	wt	MSS
763	wt	MSI-H	1629	wt	ND
789	wt	MSS	1634	wt	ND
809	wt	MSS	1666	wt	MSS
812	wt	MSS	1667	wt	ND
813	wt	MSI-H	1678	wt	MSI-L
814	wt	MSI-H	1699	wt	MSI-H
815	wt	MSI-H	1703	wt	MSI-H
825	wt	MSI-H	1705	wt	ND
869	wt	MSI-H	1708	wt	MSS
871	wt	MSI-H	1716	wt	MSS
889	wt	MSI-H	1724	wt	ND
906	wt	MSS	1733	wt	MSS
932	wt	MSI-H	1804	wt	ND
949	wt	MSI-H	1819	wt	MSI-L
968	wt	MSI-H	1851	wt	MSS
982	wt	MSI-H	1938	wt	ND
1020	wt	ND	1942	wt	ND
1052	wt	MSI-H	1989	wt	ND
1311	wt	ND	2050	wt	ND
1323	wt	MSS	2052	wt	ND
1338	wt	MSS	2145	wt	MSS
1382	wt	MSI-H	female	wt	control

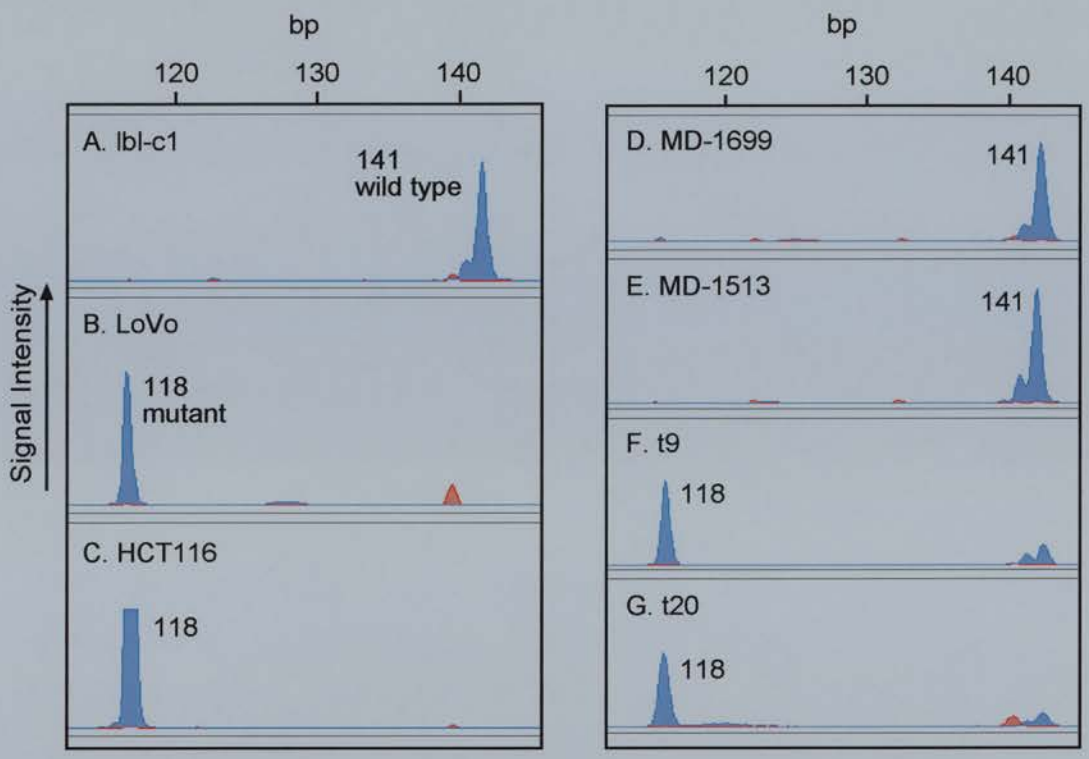


Figure 7.2 Validation of *TGFBR2* assay to detect 1bp deletions in the poly(A)$_{10}$ repeat. ABI 310 traces show peaks corresponding to the presence or absence of mutant template DNA as revealed by this assay. (A) Only the wild type peak is detected in control cell line Ibl-c1 after *HINF1* digestion. (B and C) LoVo and HCT116 are cancer cell lines with known 1bp deletions at the *TGFBR2* poly(A)$_{10}$ repeat (see text), and accordingly only mutant peaks of 118bp are detected after *HINF1* digestion. D) and E) represent two of the assays performed on constitutional DNA from patients with CRC and show only wild type peaks. The colorectal tumour from patient MD-1699 has been demonstrated previously to be MSI-H and tumour from patient MD-1513 has been demonstrated to be MSI-L. F) and G) represent two of the assays carried out on DNA derived from the tumour material of CRC patients with MSI$^+$ tumours. Mutant peaks indicate that these tumours have 1bp deletions of the poly(A)$_{10}$ tract in *TGFBR2*. A residual fraction of the digestion products were wild type and these likely represent contamination of the tumour material with normal tissue. This would suggest that the tumour cells have homozygous *TGFBR2* mutations.

7.3.2 Excess mutations arise in the poly(A)$_{10}$ tract of *TGFBR2* in lbl-1261

The *TGFBR2* assay was then utilised to analyse DNA from 1bl-1260 and lbl-1261 to address whether the mutator phenotype is associated with an excess mutation frequency in the poly(A)$_{10}$ tract of *TGFBR2* (Figure 7.3). Appreciable levels of mutant alleles were detected in DNA from lbl-1261 as indicated by the detection of a mutant peak in addition to that corresponding to wild type DNA in the undiluted sample (Figure 7.3A). The wild type peak was more prominent, suggesting that mutant alleles were present at a lower frequency in the template DNA. Intriguingly the frequency of mutant alleles indicated for lbl-1260 was not significantly different to that of wild type control DNAs (Figure 7.3B).

This analysis reveals that 1bp deletions of the poly(A)$_{10}$ tract of *TGFBR2* arise in cells derived from normal tissue but which are defective for MMR. Heterogeneity in mutation frequency between the two MMR deficient cell lines with different MMR mutations was also observed at microsatellite sequences in Chapter 4. Lbl-1261 displayed increased mutation frequency compared to lbl-1260 at the repeat markers BAT-40 and D2S123. This phenomenon also seems apparent within coding sequence, since the poly(A)$_{10}$ repeat in *TGFBR2* in lbl-1261 demonstrates an elevated level of mutation compared to lbl-1260.

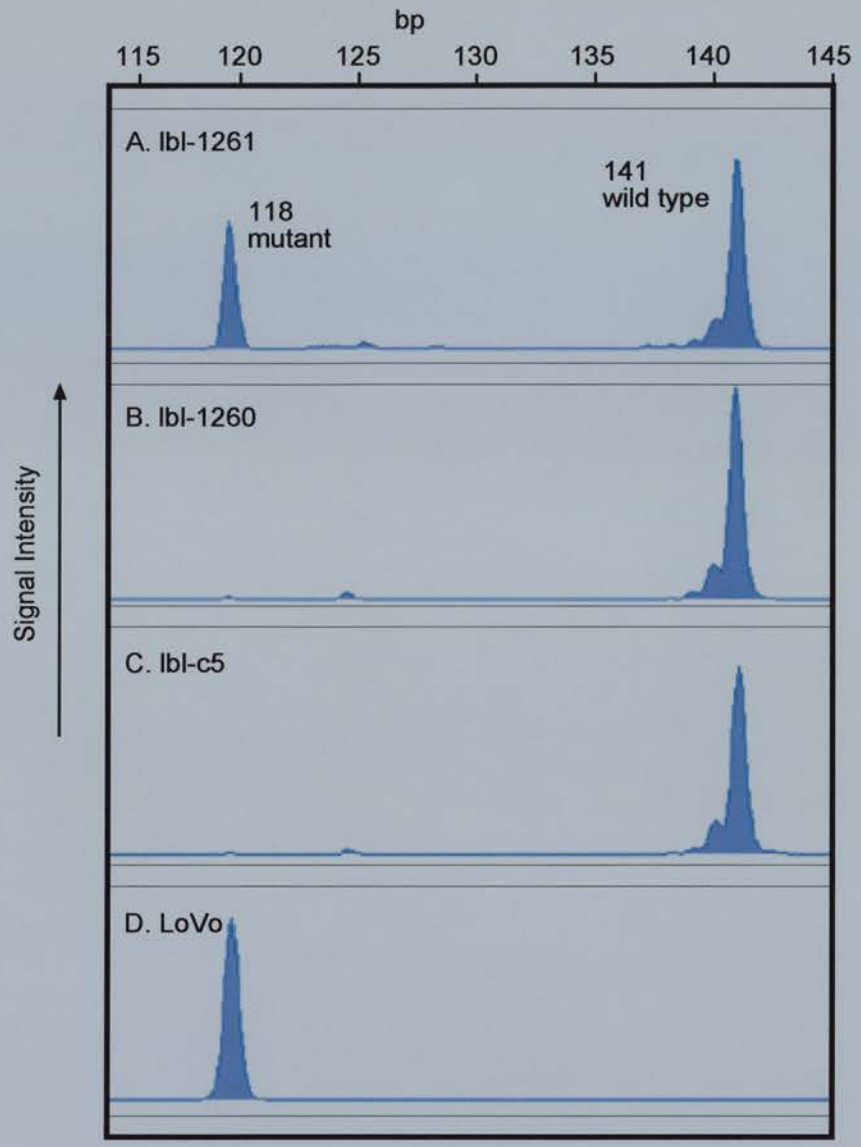


Figure 7.3 Analysis of MMR deficient cell lines Ibl-1260 and Ibl-1261 for 1bp deletions within the *TGFBR2* poly(A)$_{10}$ repeat. Representative assays of the poly(A)$_{10}$ tract in exon 3 of the *TGFBR2* gene from wild type (Ibl-c5) (C) and mutant (LoVo) (D) control cell lines showing ABI 310 traces corresponding to the presence or absence of mutant template DNA. (B) The TGFBR2 assay was performed on Ibl-1260 DNA revealing only a wild type peak. (A) A mutant peak is clearly visible for Ibl-1261.

7.3.3 Mutational heterogeneity at the *TGFBR2* poly(A)$_{10}$ tract in lbl-1261

To further characterise mutations at the poly(A)$_{10}$ repeat detected in lbl-1261 by the restriction digest assay described above, exon 3 of *TGFBR2* was PCR cloned from lbl-1261 DNA. This region was also cloned from wild type control cell line lbl-c5. Individual alleles were then sequenced across the poly(A)$_{10}$ repeat (Figure 7.4, Table 7.3).

Of 56 lbl-1261 clones sequenced, 17 (30.3%) were mutated at the poly(A)$_{10}$ tract. In contrast only 3 (6.8%) of the 44 wild type control clones were mutant. Therefore mutation at the poly(A)$_{10}$ region of *TGFBR2* was significantly increased in lbl-1261 relative to a control cell line (χ^2=8.5, p=0.003). Most of the exon 3 poly(A)$_{10}$ mutations detected in lbl-1261 were 1bp deletions (82%), supporting the validity of the results from the digestion assay described above. However a small number of 1bp insertions plus an A→G transition within the poly(A) tract were also identified (Figure 7.4 and 7.5). The frequency and the spectrum of these mutations suggest they are not constitutional variants but arose somatically. Unfortunately tissue was not available from the colorectal tumours of patient lbl-1261 to establish whether these also harboured mutations in the *TGFBR2* poly(A)$_{10}$ tract. However, the high frequency of *TGFBR2* mutations observed in the analysis of the MSI$^+$ tumours, indicates that this is highly likely.

These data provide compelling evidence that de-novo mutations can arise frequently due to inherent instability of the poly(A)$_{10}$ repeat in *TGFBR2* but these are normally repaired by a proficient DNA MMR apparatus.

Table 7.3 Cloning analysis of the poly(A)$_{10}$ repeat in *TGFBR2*. Individual alleles were sequenced from 56 lbl-1261 and 44 lbl-c5 clones. The frequency and nature of mutations detected at this 10bp repeat from nt709 to nt718 of *TGFBR2* are detailed. A mutation frequency is given by dividing the total number of mutations detected by the total amount of sequence analysed at the poly(A)$_{10}$ locus.

	Cell line	
	lbl-1261	**lbl-c5**
Total sequence analysed/Kb	0.56	0.44
Mutations identified	14 x 1bp deletion 2 x 1bp insertion 1 x a-g transition	2 x 1bp deletion 1 x 2bp deletion
Mutation Frequency, mutations/Kb	30.36	6.82

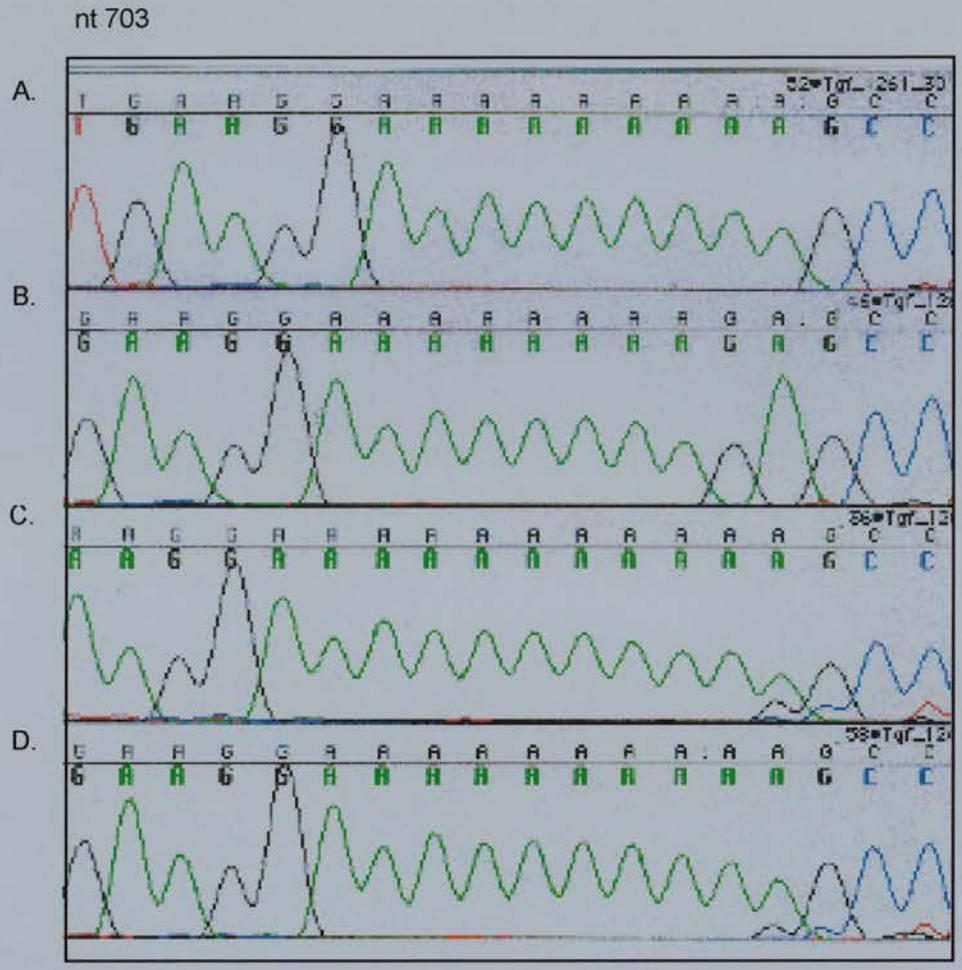


Figure 7.4 Representative sequence chromatograms of the poly(A)$_{10}$ repeat of *TGFBR2* in individual alleles cloned from Ibl-1261 template DNA. Along with wild type sequences (D), alleles were identified with 1bp deletions (A), an a to g transition (B), and 1bp insertions (C).

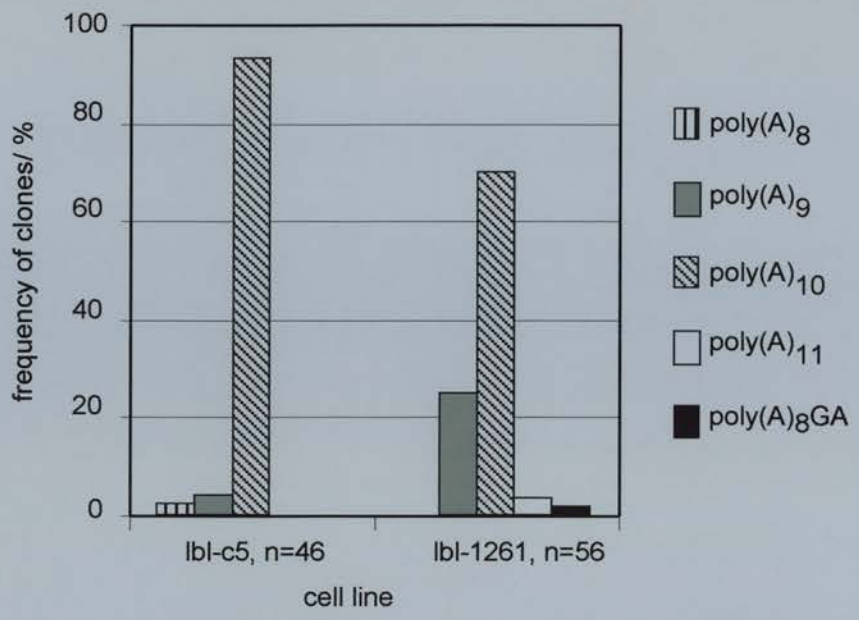


Figure 7.5 Summary of mutation frequency and spectrum at the poly(A)$_{10}$ tract of *TGFBR2*, in alleles cloned and sequenced from cell lines lbl-1261 and lbl-c5. The greater propotion of mutant alleles in lbl-1261 was statistically significant (p=0.003).

7.3.4 Determination of the relative susceptibility of regions within *TGFBR2* to mutation

The relative susceptibility to mutation of regions within the *TGFBR2* gene was determined. In particular to address whether the mutator phenotype results in predilection for mutation at the poly(A)$_{10}$ repeat sequence or whether the observations were simply the result of non-specific increase in mutation rate. Mutant allele counts arising at non-repeat regions of the *TGFBR2* gene in lbl-1261 were compared to those arising in the poly(A)$_{10}$ tract.

A 216bp region of exon 3 flanking the repetitive poly(A)$_{10}$ region and a 245bp non-repeat region of exon 4 were screened for mutations by PCR cloning and sequencing of individual alleles. Exon 4 contains a site found to be mutated in MSS colorectal cancer (Lu *et al.*, 1998) but the region lacks long repeats. No mutations were identified in this region of exon 4 in a total of 7.840kb of sequence from 32 lbl-1261 clones sequenced (Table 7.4). However, in the non-repetitive region of exon 3, 6 transition mutations were identified in a total of 10.928kb of sequence analysed from 49 lbl-1261 clones, a frequency of 0.55 mutations/kb (Figures 7.6 and 7.7 and Table 7.4). There were no mutations in the non-repetitive region of exon 3 in a total of 8.897kb from 40 control (lbl-c5) clones (Figure 7.6 and Table 7.4).

These data indicate that exon 3 of the *TGFBR2* gene appears prone to mutation, compared with exon 4 of the same gene ($p=0.045$), and also compared to the same region in a control cell line ($p=0.036$). Nonetheless, the proportion of mutations detected in the non-repeat region of *TGFBR2* is significantly lower than that for the poly(A)$_{10}$ tract ($p<0.001$) (Tables 7.3 and 7.4).

Taken together, these results suggest that exon 3 of *TGFBR2* is inherently mutable and that predilection for instability at the poly(A)$_{10}$ tract contributes to the frequent observation of *TGFBR2* mutations in MSI$^+$ tumours.

Table 7.4 Cloning analysis of two non-repeat regions in *TGFBR2*. The non-repeat region in exon 3 analysed for mutations, comprised a 216bp region excluding the $(A)_{10}$ repeat, from nt599 -10 to nt789 +26. The region analysed in exon 4 comprised a 245bp region from nt1021 to nt1266. Mutation frequency is given by dividing the total number of mutations detected, by the total amount of sequence analysed. Sequencing of individual alleles was carried out on MMR deficient cell line lbl-1261 and also for exon 3, MMR proficient control cell line lbl-c5.

	***TGFBR2* exon 3 non-repeat region**		***TGFBR2* exon 4**	
Cell lines	lbl-1261	lbl-c5	lbl-1261	lbl-c5
Total sequence analysed/Kb	11.016	8.967	7.840	ND
Mutations identified	t-c nt744 a-g nt602 a-g nt677 a-g nt702 t-c nt725 c-t nt651	None	None	ND
Mutation frequency, Mutations/Kb	0.55	Undetectable	Undetectable	ND

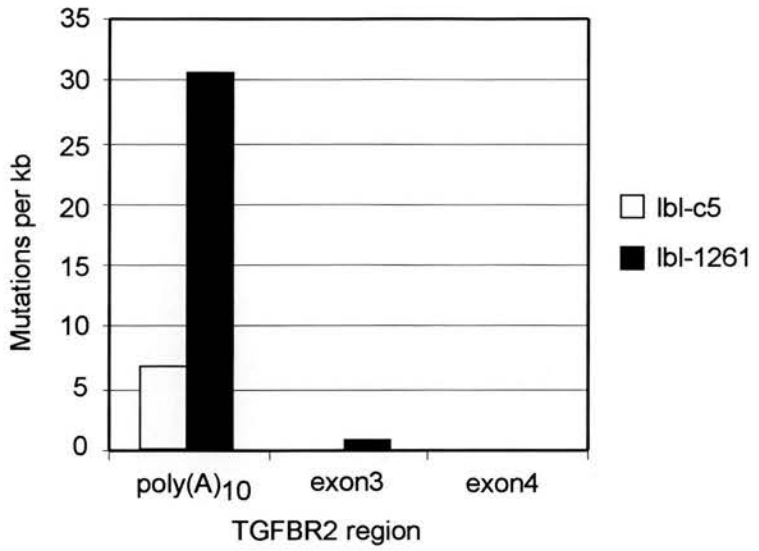


Figure 7.6 Mutational analysis of the *TGFBR2* gene comparing non-repetitive coding sequence to the poly(A)$_{10}$ tract. The non-repetitive sequence of exon 3 was analysed in lbl-1261 and lbl-c5. Lbl-1261 sequence data for a second 245 bp control non-repeat region of *TGFBR2* is also presented. Mutation data are represented as mutations per kb sequenced, to account for variation in the length of different regions analysed.

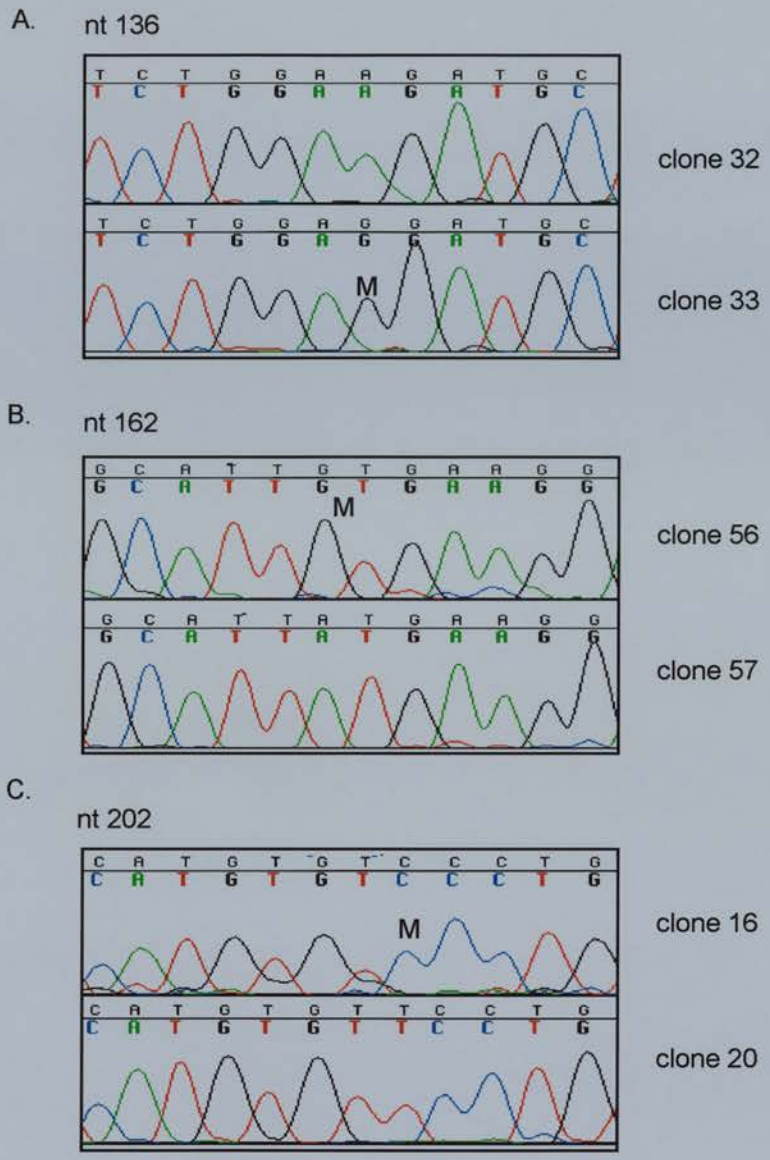


Figure 7.7 Representative mutations (M), identified in the flanking sequence around the poly(A)$_{10}$ tract of *TGFBR2* in alleles cloned from Ibl-1261 DNA. In clone 33 (A) an a to g transition was identified at nt 142. In clone 56 (B) an a to g transition was also identified at position 167 and in clone 16 (C), a t to c transition was identified at position 209. Clones 32 (A), 57 (B) and 20 (C) provide examples of alleles with wild type sequence, detected at each region.

7.3.5 Analysis of inherent instability at the BAX gene in MMR deficient cells

Since cell line lbl-1261 was particularly unstable at *TGFBR2* and non-transcribed microsatellite repeats, it was analysed for evidence that the mutator phenotype is associated with a high level of mutation in the poly$(G)_8$ repeat of the pro-apoptotic gene *BAX*. Analysis was limited to lbl-1261 in view of the more extreme phenotype at BAT-40, D2S123 and *TGFBR2*. SP-PCR was employed to genotype alleles using LoVo as a positive control (Figure 7.8). Of 164 SP-PCR products only 2 mutants (1.2%) were detected by DHPLC of the 94bp products (Figure 7.8D). Sequencing of one of these mutant SP-PCR products confirmed a 1bp insert at the poly$(G)_8$ tract (Figure 7.8E). No mutations were detected in 106 SP-PCR products analysed from a control cell line. These data suggest that mutations do occur at the repetitive poly$(G)_8$ tract of the *BAX* gene, but the frequency is much lower than that arising in *TGFBR2* and is below the level of reliable detection.

Table 7.5 Summary of SP-PCR analysis of the poly$(G)_8$ repeat in the *BAX* gene. SP-PCR products were analysed for mutations by DHPLC using a Transgenomic Wave™. Analysis was carried out on MMR deficient cell line lbl-1261 and MMR proficient cell line lbl-c5

Cell line	SP-PCR products analysed	Total mutant SP-PCR products (frequency)
lbl-1261	164	2 (1.2)
lbl-c5	106	0 (0)

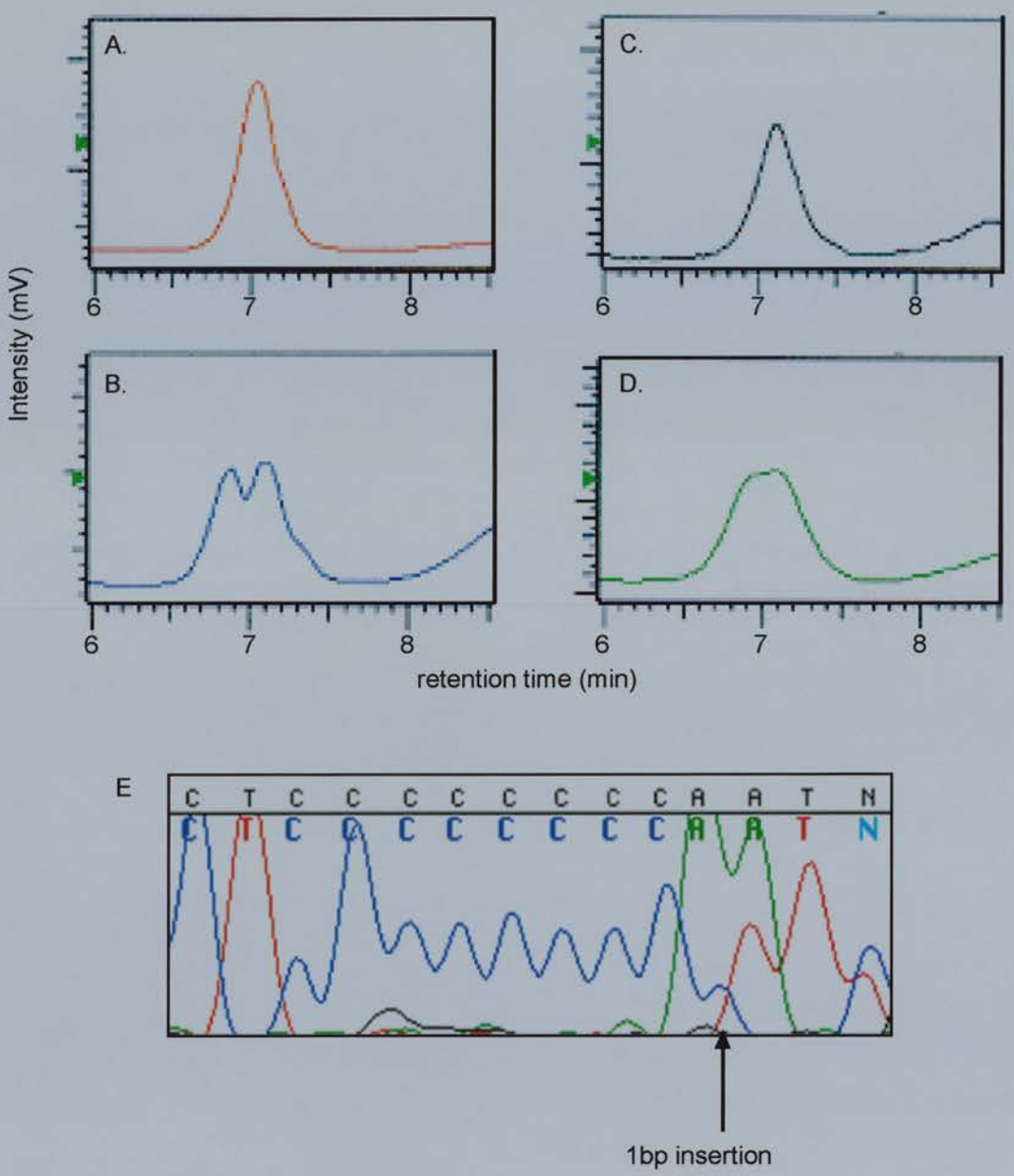


Figure 7.8 Analysis of the poly(G)$_8$ tract of the *BAX* gene in lbl-1261 DNA by SP-PCR and DHPLC. (A) DHPLC trace of the *BAX* poly(G)$_8$ region amplified from undiluted lbl-1261 DNA shows single homoduplex peak corresponding to wild type template. (B) DHPLC trace of this region amplified from cancer cell line LoVo, displays an additional heteroduplex peak corresponding to the presence of a known mutation in this cell line (see text). (C) This SP-PCR product , amplified from lbl-1261 DNA has only a wild type homoduplex DHPLC peak. However, the DHPLC trace of the SP-PCR in (D), shows a flattened peak indicative of the presence of homoduplex and heteroduplex products corresponding to a mutant template. (E) Sequencing of SP-PCR product shown in D) reveals the presence of a heteroduplex product containing an allele with a 1bp insertion in the poly(G)$_8$ tract. The reverse sequence is shown.

7.4 Discussion

The data presented in this chapter demonstrate a substantial level of inherent instability at coding sequences, as a consequence of MMR deficiency in cell line lbl-1261 that is derived from normal tissue. These studies have provided evidence that inherent mutability in the *TGFBR2* gene contributes to the observation of similar mutations in MSI^+ tumours. The instability shows a clear propensity to the same $poly(A)_{10}$ tract that is frequently mutated in MSI^+ tumours. The analysis of the $poly(G)_8$ tract of the *BAX* gene indicates that there are varying degrees of inherent mutational instability within coding regions known to have a high frequency of mutation in tumours, and suggests differential contributions from mutational mechanisms and the effects of the cancer phenotype itself.

Two independent experimental approaches were employed using the MMR deficient cell line, lbl-1261, to demonstrate that the $poly(A)_{10}$ tract of *TGFBR2* is intrinsically hypermutable. These data show that mutations can be detected as a consequence of MMR deficiency when the effects of selection pressure and other confounding molecular variables normally present in tumour cells are minimised.

These studies were performed in cell lines that represent a model assay system and as such, selection cannot be entirely negated. There is evidence that EBV transformed lymphoblast cell lines are resistant to the effects of TGFB1 due to selection against TGFBR2 expression (Inman and Allday, 2000). However, this effect is unlikely to bias the frequency of mutations at the $poly(A)_{10}$ tract for a number of reasons, 1) Cells used in this study were not treated with TGFB1 and thus were not under heavy selection as in the previous study by Inman *et al* (Inman and Allday, 2000), 2) Reduced expression of TGFBR2 in EBV transformed cell lines has been shown definitively <u>not</u> to be due to mutation of the $poly(A)_{10}$ repeat in exon 3 (Inman and Allday, 2000), 3) The fact that the reduced TGFBR2 expression previously observed is due to a mechanism other than mutation, underscores the fact that selection does not contribute appreciably to the accumulation of inactivating mutations, 4) The comparison with EBV transformed control lines subject to the

same conditions in culture, shows conclusively that there is an important MMR dependent effect.

Since a number of different *TGFBR2* mutant alleles were detected, the findings are also consistent with the occurrence of somatic events rather than pre-existing constitutional heterozygous mutations or mosaicism. One bp deletions were the most common mutation which is in line with the pattern of mutations observed at this locus in MSI^+ tumours (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b). Finally, the *TGRBR2* poly$(A)_{10}$ mutations identified in cells derived from normal tissue, are unlikely to reflect contamination from primary tumour DNA circulating in the blood. These cell lines were specifically tested for the presence of epithelial cells in the FACS analysis presented in Chapter 3. None of the cells in lbl-1261 were found to express an epithelial cell specific marker, indicating that this cell line is not contaminated with micro-metastasising cells. Furthermore, despite the fact that *TGFBR2* mutations were detected in the tumours of MD-1699 and MD-528 (Table 7.1), no mutations were identified in the matched normal DNA from these individuals (Table 7.2) during analysis of matched normal and tumour DNA from other CRC patients (7.3.1).

The fact that mutations in sequences encoding tumour suppressor genes were detected in cells derived from a normal B-cell lineage is remarkable in itself. That lbl-1261 appears particularly prone to mutation may be due to specific influences from MMR genes. Indeed, a recent report (Vilkki *et al.*, 2001) did not detect any *TGFBR2* mutations in non-cancerous autopsy material in a child with homozygous *MLH1* mutations. Interestingly, cell line lbl-1260, which contains a heterozygous *MLH1* mutation, displays a lower incidence of mutation at both coding and non-transcribed SSRs (Chapter 4). The fact that excess mutations were not observed at the poly$(A)_{10}$ tract in lbl-1260 may indicate a threshold effect of MMR activity in the cell lines studied here, which has an influence on the ability for replication errors to be repaired. Levels of MMR activity on a biochemical level in these cell lines has been previously investigated with no discernible difference in repair activity of nuclear extracts from lbl-1261 and lbl-1260 (Parsons *et al.*, 1995a). However the frequency of mutations measured at microsatellite loci was significantly elevated in lbl-1261 compared to lbl-1260 observed both in this study (Chapter 4) and

previously (Parsons *et al.*, 1995a). This may reflect subtle differences in repair capabilities in these two cell lines resulting from the different germline MMR mutations that they harbour. A model implying that the threshold of repair activity directly influences the spectrum of mutations observed has been recently proposed (Cahill *et al.*, 1999). Further analysis of low level MMR activity in these MMR deficient cell lines would be merited to address if this hypothesis can account for the mutational differences between lbl-1261 and lbl-1260 at the *TGFBR2* gene.

Mutations revealed by cloning analysis in a control cell line may indicate a background error rate or may be genuine mutations occurring at a level undetectable by the restriction digest analysis. Indeed there is little investigation into low-level mutation frequencies in cancer causing genes in either normal tissue or in control cell lines. However, both analyses demonstrate a statistically significant excess of mutations in lbl-1261 compared to a control cell line.

Few studies address the mutation frequency at the non-repetitive sequences surrounding the mutable repeat tracts in genes implicated in tumourigenesis of MSI^+ tumours (Takenoshita *et al.*, 1997). In this chapter it has been specifically demonstrated that non-repeat regions of *TGFBR2* are subject to levels of instability that are statistically significantly lower than at the poly$(A)_{10}$ repeat. These data suggest that it is the repeat sequence and not the entire *TGFBR2* gene, which is prone to instability consequent of MMR deficiency.

Analysis of the poly$(G)_8$ repeat region in the *BAX* gene, suggests that although this is mutated frequently in MMR deficient tumours, it does not share the same propensity for mutation observed at the poly$(A)_{10}$ repeat of *TGFBR2*. These results give considerable insight into susceptibility of the *BAX* and *TGFBR2* genes to mutations rising exclusively as a result of MMR defects. The high mutability of the poly$(A)_{10}$ tract in *TGFBR2* is likely to be a consequence of increased replication error at this region. Such error is well documented and is dependent on the number of repeats in the tract (Parsons *et al.*, 1995a; Dietmaier *et al.*, 1997; Sturzeneker *et al.*, 2000). Chromatin structure within or surrounding such repetitive tracts has also been suggested to contribute to mutability (Zhang *et al.*, 2001). In this regard it is interesting to note that a low frequency of mutation was observed in the flanking

sequence surrounding the poly(A)$_{10}$ tract in lbl-1261, that was significantly greater than that observed for a region of exon 4 of the same gene possibly indicating that chromatin structure may play a role in the mutability of the poly(A)$_{10}$ tract.

That the *TGFBR2* poly(A)$_{10}$ repeat is so readily prone to mutation in the absence of MMR goes some way to explain why these mutations are observed with such a consistently high frequency within MMR deficient tumours (Markowitz *et al.*, 1995; Markowitz, 2000) and also in early adenomas (Grady *et al.*, 1998; Abdel-Rahman *et al.*, 1999). Taken together with previous studies, the data presented here supports the notion that MMR inactivation occurs very early in neoplastic transformation and results in accumulation of mutations in the *TGFBR2* poly(A)$_{10}$ tract. Subsequently, selection pressure for homozygous mutations would be anticipated to result in clonal selection within the tumour.

Since only a very low level of *BAX* mutations were detected in the most unstable cell line, lbl-1261, this suggests that *BAX* is inherently more stable and that selection pressure may play a greater role in generating poly(G)$_8$ mutations, than is the case for *TGFBR2* mutations. The data presented here suggests that *BAX* mutations arise infrequently and this may explain why they are less frequently observed, and are not detectable at such an early stage in tumourigenesis as TGFBR2 mutations (Rampino *et al.*, 1997; Yagi *et al.*, 1998; Abdel-Rahman *et al.*, 1999)

These findings indicate that intrinsic mutational instability is an important determinant of mutation frequency observed in MSI$^+$ tumours. Use of cells derived from normal tissue has allowed analysis of the mutation frequency consequent upon MMR deficiency, while minimising bias from the effects of selection and confounding abnormalities. It will be of interest to further these studies and determine which other coding gene sequences are similarly susceptible to mutation in cell lines lbl-1260 and lbl-1261 as this might identify important genes involved in MMR dependent tumour initiation and progression.

Chapter 8

Analysis of the Susceptibility of β-catenin/TCF Pathway Genes to Mutation in MMR Deficient Cells Derived from Normal Tissue

8.1 Introduction

With the exception of *TGFBR2* and *BAX*, there is conflicting evidence on the repertoire of genes that when altered, contribute directly to the oncogenic properties of MSI^+ tumours. In addition, it is not clear whether such alterations arise as a consequence of genomic instability or selection.

Abrogation of the β-catenin/TCF signalling pathway is an important step in the development of the majority of CRCs (Miyoshi *et al.*, 1992; Powell *et al.*, 1992). In particular, the gatekeeping role of the *APC* gene in CRC is well established (Kinzler and Vogelstein, 1996; Fearnhead *et al.*, 2001). However, there are conflicting data on the contribution of *APC* mutations to the initiation and progression of MSI^+ tumours as discussed in Chapter 1. Whether *APC* mutations are associated with MMR defects has not been definitively established (Heinen *et al.*, 1995; Huang *et al.*, 1996; Konishi *et al.*, 1996; Olschwang *et al.*, 1997; Homfray *et al.*, 1998). Where a high frequency of *APC* mutations have been reported in MSI^+ tumours, it is unclear as to whether they arose prior to the loss of MMR or as a consequence of such defects (Huang *et al.*, 1996; Homfray *et al.*, 1998). The conflicting data provide little resolution of the selection vs mutation debate (Huang *et al.*, 1996; Homfray *et al.*, 1998; Tomlinson and Bodmer, 1999; Loeb, 2001).

Activating mutations in the regulatory domain of β-catenin, have also been observed in some MSS colorectal tumours that are wild type for *APC* (Morin *et al.*, 1997; Iwao *et al.*, 1998; Sparks *et al.*, 1998). However, it is unclear as to how frequently *CTNNB1* mutations substitute for *APC* mutations in MSI^+ CRCs (Sparks

et al., 1998; Mirabelli-Primdahl *et al.*, 1999; Miyaki *et al.*, 1999; Salahshor *et al.*, 1999). Again these contradictory observations make it difficult to conclude whether or not mutational abrogation of the β-catenin/TCF pathway is an important event in MSI^+ CRCs, albeit via alternative mechanisms to those in MSS cancers (Huang *et al.*, 1996).

In light of the conflict in the available data, the research presented in this chapter set out to investigate the effect of defective MMR on the β-catenin/TCF pathway. In particular the work addresses whether mutations in β-catenin/TCF pathway genes can arise as a consequence of defects in MMR in the absence of selection pressure. Differences in inherent susceptibility to mutation of β-catenin/TCF pathway genes may contribute to the frequency with which they are mutated in MSI^+ tumours. Therefore regions of *CTNNB1* and *APC* were analysed for mutations, in cell lines lbl-1260 and lbl-1261. It has already been demonstrated in previous chapters that since these cell lines have defects in MMR but are derived from non-cancerous tissue, the inherent instability of particular gene sequences can be effectively revealed.

The inherent stability of exon 3 of the *CTNNB1* was analysed since it encompasses the regulatory phosphorylation sites that are mutated in some MSI^+ CRCs, which lack *APC* mutations. For example, mutations at codons 32, 33, 34, 37, 41, 43, and 45 that are located within exon 3 have been previously reported (Miyaki, *et al.*, 1999; Mirabelli-Primdahl *et al.*, 1999; Akiyama *et al.*, 2000). Three regions of the *APC* gene were also analysed (Figure 8.1). The poly$(GT)_4$ repeat in exon 10 was specifically analysed since it has been previously demonstrated to show mutational mosaicism in the parent of an FAP patient (Farrington and Dunlop, 1999). The mosaicism at this repeat region was speculated to have arisen due to genetic instability (Farrington and Dunlop, 1999). In addition, two regions of exon 15 were also examined. These were B1-D2 of exon 15 that spans codons 736 to 1054 and region G1-I2 of exon 15 that spans codons 1256 to 1641. Mutations in both these regions have been identified previously in MSI^+ tumours. In particular, mutations have been detected at SSRs within these loci and these represent potential sites of mutation in the presence of MMR defects (Huang *et al.*, 1996). Such repetitive sites

include a poly(A)$_4$ at nt2822-5, a poly(A)$_5$ at nt2540-4, an (A)$_5$T(A)$_3$ at nt4363-71 and a poly(A)$_6$ at nt4664-9. Furthermore, region G1-I2 encompasses the previously reported mutation cluster region located at codons 1286-1513 (Miyoshi *et al.*, 1992).

By utilising lbl-1261 and lbl-1260 to study these coding sequence loci for susceptibility to mutation, the approach comprises an alternative to previous studies that have carried out mutational analysis of *APC* and *CTNNB1* genes in cancer cells.

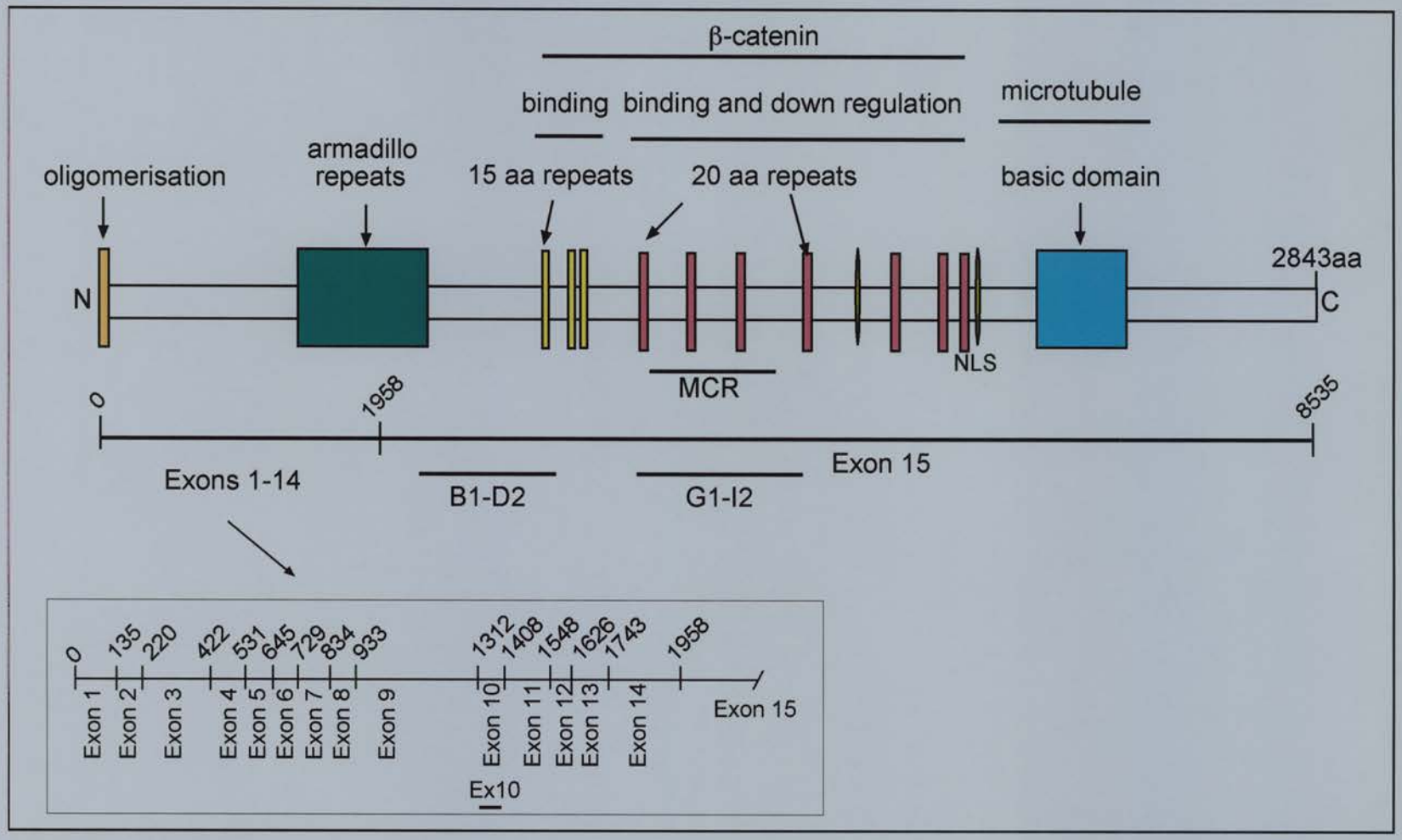


Figure 8.1 Schematic diagram of APC protein and gene. The genomic structure and various protein motifs are illustated. The mutation cluster region (MCR) is located in the central domain of the protein. In this chapter the entirity of exon 10 and two regions of exon 15 were analysed for evidence of inherent instability. They are indicated as Ex10, B1-D2 and G1-I2 respectively.

8.2 Methodological Overview

8.2.1 PCR of *CTNNB1* and *APC* fragments

A 251bp region from nt14-12 to nt241+12 that encompasses exon 3, was amplified from the *CTNNB1* gene using primers CTNNB1ex3F and CTNNB1ex3R (2.4.1) and 100ng of cell line DNA template as described in 2.4.2. A 955bp region of exon 15 of the *APC* gene (B1-D2) from nt2206 to nt3161, was amplified using primers APCex15B1 and APCex15D2 (2.4.1) and 100ng template DNA as described in 2.4.2. A 1156bp region of exon 15 of the APC gene (G1-I2) from nt3767 to nt4923 was also amplified using primers APCex15G1 and APCex15I2 (2.4.1) as described in 2.4.2.

The Expand High Fidelity PCR system was used in all amplifications to allow the most faithful replication of DNA templates.

8.2.2 Cloning of individual *CTNNB1* and *APC* alleles

Individual alleles were cloned from PCR products described in 8.2.1 into TA cloning vectors as described in 2.6.1. All clones were checked for the presence of an insert of appropriate size by direct PCR analysis from each bacterial colony using M13F and M13R primers as described in 2.4.2. Bacteria were grown and plasmid DNA harvested as described in 2.6.2 and 2.2.4.

8.2.3 Sequencing of cloned alleles

Individual alleles were sequenced from plasmid DNA as described in 2.7. Each insert was sequenced using M13F and M13R vector specific primers (2.4.1). Sequence data was then analysed as described in 2.7.5. For *APC* exon 15 region B1-D2, nt2229-3139 was specifically analysed for mutations and for *APC* exon 15 region G1-I2, nt3791-4905 was specifically analysed for mutations.

8.2.4 Estimation of PCR error

Observation of statistically significant differences in mutation frequency at a given locus between alternative cell lines strongly suggests that PCR error does not significantly bias the results obtained. In addition, observations of statistically significant differences between the same cell lines at alternative loci are difficult to reconcile with PCR error. However, to definitively rule out introduction of error from Taq polymerase as a valid explanation of the results, the expected number of PCR errors was specifically estimated. The number of polymerase errors expected within the total amount of sequence analysed (nt) for alleles cloned from a given cell line was calculated from the error rate of the Expand Taq polymerase, according to the information provided with the Expand High Fidelity PCR system (Boheringer Mannheim) (Table 8.1).

Table 8.1 Expected errors from Expand Taq polymerase. (Boheringer Mannheim) Expand Taq Polymerase has an error rate of 8.5×10^{-6}. The number of PCR fragments with errors is expressed in % and depends on fragment length and cycle numbers. * Amplifying up to 3kb, the effective cycle number is around 60% of the machine cycles, e.g.: 30-programmed cycles correspond to 18 effective cycles. Table adapted from Table 1 of the Expand product information sheet (Boheringer Mannheim).

PCR Fragment size [kb]	**PCR fragments with errors [%]**		
	Number of effective* cycles		
	10	**20**	**30**
0.2	1	2	2
1.0	4	8	12

In all PCR reactions for which the number of errors was calculated, amplification was for 35 machine cycles (21 effective cycles).

The effective number of fragments analysed for each sample was calculated according to;

$$\frac{\text{Total amount of sequence analysed per sample}}{\text{Individual fragment length}} = \text{Effective number of fragments}$$

The expected number of errors per sample was then calculated from the percentage of fragments with an error as calculated using the table above. Malcolm Sheppard, Technology Support Supervisor at Roche Diagnostics, verified these calculations.

8.2.5 SP-PCR of *APC* exon 10

The PCR fragment encompassing the poly(GT)$_4$ repeat in exon 10 of *APC* was amenable to SP-PCR/Transgenomic Wave™ analysis and thus a 210bp region from nt1313 –59 to nt1408 +55 encompassing exon 10 of the *APC* gene was amplified using the SP-PCR technique described in 2.4.4 and 2.4.5. Primers APCex10F and APCex10R were used for amplification as detailed in 2.4.1.

8.2.6 DHPLC of *APC* exon 10 using a Transgenomic Wave™

Analysis of *APC* exon 10 SP-PCR products was carried out by DHPLC using a Transgenomic Wave™ as described in 2.5.2 using the *APC* exon 10 melting gradient (2.5.3)

8.2.7 Statistical analysis

Relative mutation frequency was calculated at each locus for each sample analysed. The total number of mutations observed was divided by the total amount of high quality sequence analysed in kilobases to give a mutation frequency/kb. Sequencing with forward and reverse, vector specific primers gave high quality sequence across all the *CTNNB1* alleles included in this analysis. For inserts cloned from the two *APC* exon 15 regions partial read-through of alleles occurred in some instances since the inserts were longer. As this was observed to be a stochastic event

with no apparent bias for alleles from any one sample, all good quality sequence data was included for analysis.

It was assumed that a mutation might be equally likely to occur at any individual base pair by chance. Differences in the number of nucleotides mutated out of the total nucleotides analysed at any given locus from different DNA templates were then evaluated for significance using a Chi-Squared analysis as described in 2.10.1. Where observed counts were less than 5, a Fishers Exact test was employed as described in 2.10.1

Chi-Squared or Fishers Exact tests were also employed appropriately (2.10.1) to test whether the number of mutations observed in any given sample, differed significantly to that expected from PCR error.

Significant differences in any given mutation occurring out of the total number of mutations detected between different samples was evaluated by a Chi-Squared test (2.10.1) or again, a Fishers Exact test if expected counts were less than 5.0.

8.3 Results

8.3.1 MMR defects are not associated with excess mutations in the regulatory domain of the *CTNNB1* gene

Mutations arise at phosphorylation sites within exon 3 of the *CTNNB1* gene in CRCs that do not have *APC* mutations (Sparks *et al.*, 1998; Morin *et al.*, 1997; Mirabelli-Primdahl *et al.*, 1999; Miyaki *et al.*, 1999). To address whether such mutations can arise exclusively as a result of MMR defects without the selection bias that occurs during tumourigenesis, the level of spontaneous mitotic mutations was assessed in exon 3, in MMR defective cell line lbl-1261, which is derived from normal tissue. Lbl-1261 was the focus of this analysis because of the high level of instability demonstrated at both non-coding and coding SSRs in experiments presented in previous chapters.

The cloning strategy used in the previous analysis of the *TGFBR2* gene (Chapter 7), effectively detected both frameshift and mis-sense changes. Therefore, a similar cloning strategy was employed to assess the mutational status of *CTNNB1* in individual alleles. A 251bp region encompassing exon 3 of *CTNNB1* was PCR cloned from lbl-1261 and from MMR proficient lymphoblastoid cell line lbl-a. Sequencing of 89 alleles (22,339 bp) from lbl-1261 template DNA revealed 2 mutations in exon 3, a mutation frequency of 0.090 mutations/kb (Table 8.2). Both mutations identified were T→C transitions but did not occur on the serine/threonine phosphorylation sites previously reported as being hotspots for mutation in colorectal tumours (Table 8.3) (Kitaeva *et al.*, 1997; Sparks *et al.*, 1998; Mirabelli-Primdahl *et al.*, 1999; Miyaki *et al.*, 1999; Akiyama *et al.*, 2000). Sequencing of 74 alleles (18,574bp) cloned from lbl-a template DNA, also revealed two transition mutations, a mutation frequency of 0.108 mutations/kb (Table 8.2). Again these did not occur at previously described mutational hotspots (Table 8.3). Hence, within the levels of detectability there was no apparent difference in the frequency of mutations in *CTNNB1* alleles from lbl-1261 and the MMR proficient control cell line lbl-a ($p>0.99$). The data here indicate, that a high level of inherent mutations that are

usually repaired by MMR, do not arise within the regulatory domain of *CTNNB1*. These data strongly suggest that when selection pressure and other molecular variabilities consequent of tumourigenesis are minimised, exon 3 of *CTNNB1* is not particularly prone to mutation even when MMR is defective.

Table 8.2 Mutational analysis of *CTNNB*1 exon 3 alleles cloned from MMR deficient (lbl-1261) and MMR proficient (lbl-a) cell lines. There is no significant difference in the frequency of mutation between the two cell lines.

Cell line	No. clones	Full length clones	Missing seq/kb	Total Seq/kb	Number of Mutations	Mutation freq (Mutations/kb)
lbl-1261	89	89	0	22.339	2	0.090
lbl-a	74	74	0	18.547	2	0.108

Table 8.3 *CTNNB1* mutations identified in individual alleles cloned from lbl-a and MMR lbl-1261.

Cell line	Clone ID	*CTNNB1* change	nt	Codon	Target Sequence
lbl-1261	1	T-C	88	30	TpA
	91	T-C	141	47	TpG
lbl-a	7	C-T	90	30	CpC
	36	A-G	107	51	ApA

8.3.2 Analysis of *APC* exon 10 for inherent instability consequent of MMR defects

The poly$(GT)_4$ tract of *APC* exon 10 has been shown previously to exhibit somatic mutational mosaicism in the unaffected parent of an FAP patient and thus it was reasoned that this locus may be inherently unstable (Farrington and Dunlop, 1999). The mutational status of a 207bp region encompassing all 96bp of exon 10 of the *APC* gene was analysed by SP-PCR followed by DHPLC using a Transgenomic Wave™, from lbl-1261 and control (lbl-c5) template DNA. No mutations were identified from the $(GT)_4$ repeat in APC exon 10 in 115 SP-PCRs amplified from MMR deficient cell line lbl-1261, although a single transition mutation was identified in the 5' intronic region in one reaction (Figure 8.2). After analysis of 26 SP-PCR products from MMR proficient cell line lbl-c5 template, no mutations were identified and further assessment was abandoned.

There is no evidence from this data to suggest that the $(GT)_4$ repeat of exon 10 of the *APC* gene is particularly prone to accumulating mutations that are usually repaired by MMR.

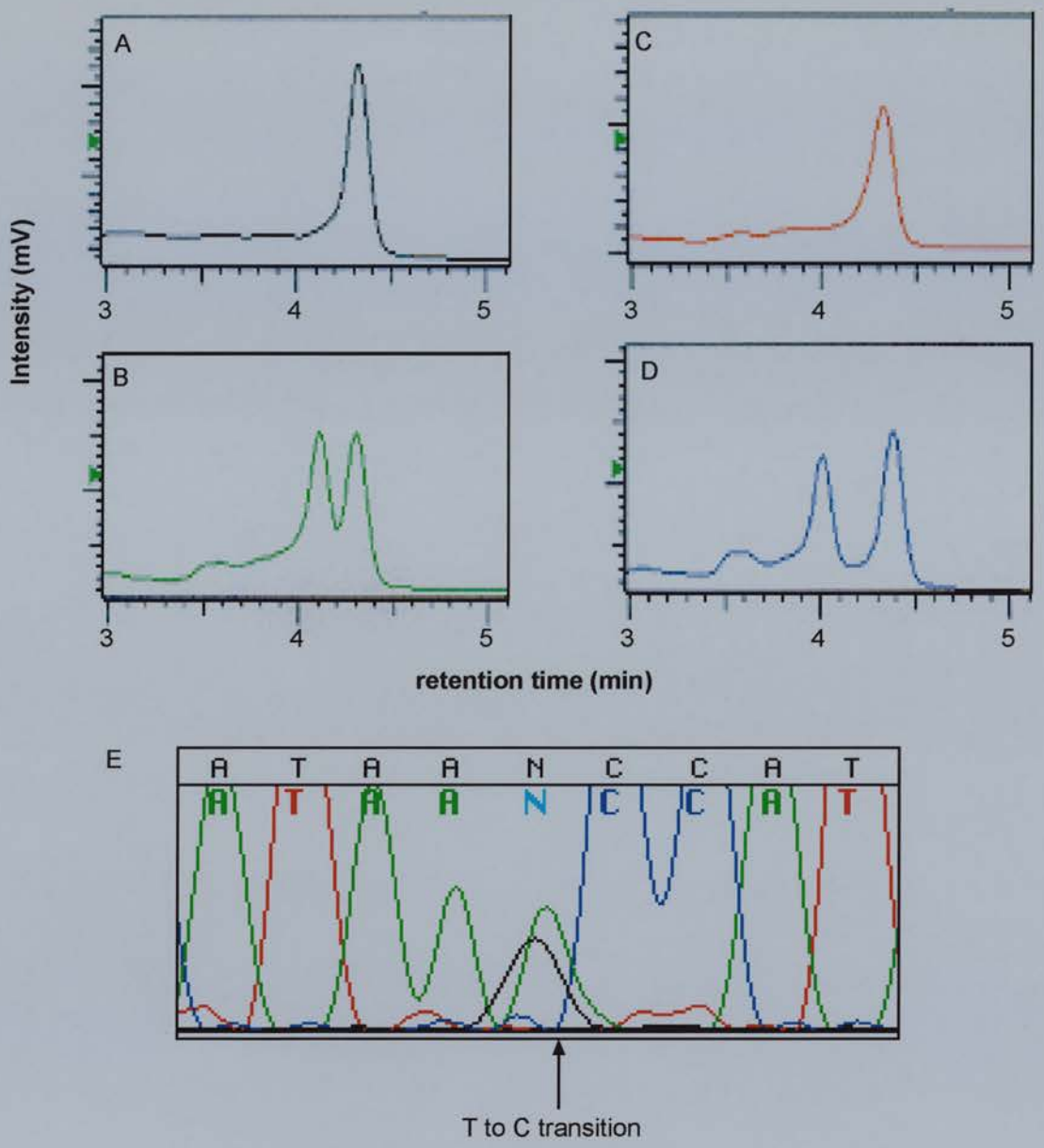


Figure 8.2 Mutational analysis of exon 10 of the *APC* gene by SP-PCR and DHPLC. (A) Exon 10 amplified from undiluted Ibl-1261 DNA shows a single homoduplex peak when analysed using a Transgenomic Wave^{Tm}. (B) The same region amplified from an FAP patient with a known mutation in the $(GT)_4$ repeat in exon 10 displays an additional heteroduplex peak. (C) Wild type SP-PCR products also show a single homoduplex peak. (D) One SP-PCR product was identified as having an additional heteroduplex peak corresponding to the presence of a mutant allele. (E) The mutant SP-PCR product was identified as having an intronic, T to C transition at nt1331 -23 in one allele, by sequencing. The reverse sequence is shown.

8.3.3 Analysis of inherent instability in exon 15 of the *APC* gene in MMR deficient and MMR proficient lymphoblastoid cell lines

In order to further address whether the *APC* gene is inherently unstable and to gain insight as to whether mutations arise in *APC* that are usually repaired by MMR defects, two further regions of the *APC* gene were analysed. Region B1-D2 of exon 15 spans codons 736 to 1054 and region G1-I2 of exon 15 spans codons 1256 to 1641 and these regions are illustrated in Figure 8.1.

Individual alleles from regions B1-D2 and G1-I2 were cloned from MMR deficient cell lines (lbl-1261 and lbl-1260) that are derived from normal tissue and from control MMR proficient lymphoblastoid cell lines (lbl-c5, lbl-a). Due to the amount of sequencing involved in these experiments, analysis of the second region G1-I2 was focused towards the highly unstable MMR deficient cell line, lbl-1261 and a MMR proficient control cell line (lbl-a). Individual alleles were sequenced and analysed for mutations from nt2229-nt3139 (B1-D2) and nt3791-nt4905 (G1-I2). A total of 80,014 bp of sequence was analysed from 92 alleles at region, B1-D2, cloned from lbl-1261 template DNA. 13 mutations were identified, a mutation frequency of 0.16 mutations/kb. (Table 8.4). A similar frequency of mutation was observed at the second region, spanning the mutation cluster region of the *APC* gene. Eleven mutations were identified in 85 G1-I2 alleles (85,090 bp) cloned from lbl-1261 DNA (Table 8.4), a mutation frequency of 0.13 mutations/kb. There is no significant difference in the level of mutation between the two regions of exon 15 analysed from lbl-1261 DNA ($\chi 2=0.313$; $p=0.576$). Furthermore, this level of mutation was not significantly different from that observed in the analysis of exon 3 of the *CTNNB1* gene ($p=0.542$), indicating that these two regions of *APC* are also not particularly prone to mutation when MMR is defective.

Table 8.4 Analysis of individual alleles cloned and sequenced from two regions of exon 15 of the APC gene, from lbl-1261, lbl-1260, lbl-c5 and lbl-a. B1-D2 = 910bp and G1-I2 =114bp.

Region	Cell line	No. clones	Full length clones	Missing seq./kb	Total seq./ kb	No. mutations	Mutation freq. (mutations /kb)
B1-D2	lbl-1261	92	77	3.706	80.014	13	0.16
	lbl-1260	98	59	11.599	77.581	54	0.69
	lbl-a	71	63	2.451	62.159	39	0.63
	lbl-c5	48	40	2.028	41.652	24	0.58
G1-I2	lbl-1261	85	65	9.600	85.090	11	0.13
	lbl-a	38	23	8.210	34.122	20	0.59

In contrast to the lack of instability in lbl-1261 at the two regions of *APC* exon 15 that were analysed, a significantly higher level of mutational instability was observed in two control lymphoblastoid cell lines (lbl-a and lbl-c5) (Figure 8.3 and Table 8.4). A total of 62,159 bp of sequence from region B1-D2 of the *APC* gene was analysed from 71 alleles cloned from lbl-a DNA. A total of 39 mutations were detected, a mutation frequency of 0.63 mutations/kb (Table 8.4). Examples of the mutations are illustrated in Figure 8.4. At the second region G1-I2, 20 mutations were detected in 34,122bp of sequence analysed from 38 alleles cloned from lbl-a DNA, (0.59 mutations/kb). There was no significant difference in the level of mutation at these two regions in lbl-a ($\chi 2=0.061$; $p=0.804$). A second control cell line (lbl-c5) was also analysed at B1-D2 and the frequency of mutations did not differ significantly from that observed for lbl-a ($\chi 2=0.108$; $p=0.743$) (Figure 8.3 and Table 8.4). These results are striking since they indicate that the MMR proficient cell lines, lbl-c5 and lbl-a, display a level of mutation that is significantly higher than that of MMR deficient cell line lbl-1261, at two regions of the *APC* gene ($\chi 2=21.59$, $p<0.001$ at B1-D2 and $\chi 2=19.55$; $p<0.001$ at G1-I2) (Figure 8.3).

Differences in mutation frequency between lbl-1260 and lbl-1261 at non-coding and coding repeat sequences have been demonstrated in experiments presented in

previous chapters. Therefore, it was not surprising that a significant difference in the frequency of mutation between lbl-1260 and lbl-1261 at region B1-D2 of the *APC* gene was also revealed, albeit with lbl-1260 displaying a higher frequency of mutation. 54 mutations were detected in 98 alleles (77,581bp) sequenced, a mutation frequency of 0.69 mutations/kb (Table 8.5). This level of mutation is not significantly different to observations in control lymphoblast cell lines ($\chi 2=0.55$, $p=0.459$) but is significantly increased compared to MMR deficient cell line lbl-1261 ($p<0.001$) (Figure 8.3).

The data presented from this extensive analysis of low level stability at two regions of the *APC* gene in MMR proficient and MMR deficient lymphoblastoid cell lines suggests three major points. Firstly, in cell line lbl-1261 there is no evidence to indicate that mutations usually repaired by MMR accumulate within the *APC* gene. The frequency of mutations detected in lbl-1261 does not differ significantly to the level detected in either exon 4 of the *TGFBR2* gene ($p=0.297$) analysed in Chapter 7 or in exon 3 of the *CTNNB1* gene ($p=0.542$). These regions were also concluded not to be particularly susceptible to mutation when MMR is defective.

Secondly, a statistically significant, four-fold difference in mutation frequency was detected in control, MMR proficient cell lines, lbl-a and lbl-c5 compared to MMR deficient cell line lbl-1261. This increased level of mutation in control lymphoblastoid cell lines is also significantly different to that observed in these cell lines in exon 3 of the *CTNNB1* gene ($p=0.002$) and also in the non-repeat region of exon 3 of *TGFBR2* ($p=0.004$). The increased mutation frequency detected at two regions of the *APC* gene in lbl-c5 and lbl-a cannot be accounted for by Taq polymerase error in the PCR step during the allele cloning protocol (Table 8.5). Twenty-nine of the mutations in the control cell lines were sequenced in duplicate and all mutations were confirmed. Approximately half of each allele was also sequenced in duplicate and no instances of discrepancy (false positives or false negatives) were observed, indicating that sequencing error was not a significant factor. In any case, the statistically significant difference in mutation frequency between the control cell lines and lbl-1261 at these regions and also between the *APC* gene and coding sequence in alternative genes in MMR proficient cell lines,

argues that biological difference rather than artefactual error contributed to these observations.

Finally, a statistically significant difference in the frequency of mutation at the *APC* gene between lbl-1260 and lbl-1261 is revealed in this analysis. This observation furthers the notion that there is a dosage dependent effect from the MMR defects that these cell lines harbour, on the frequency of mutations detected at alternative loci. Differences in underlying levels of instability in these cell lines, as evidenced at non-coding microsatellites (Chapter 4), may be important in modulating the frequency of mutation accumulation at many loci including the *APC* gene. However, it is intriguing that at the *APC* gene, mutation frequency is higher in lbl-1260 (and MMR proficient control cell lines) than in lbl-1261 in contrast to the higher instability detected in lbl-1261 at other loci studied.

Table 8.5 Expected number of errors attributable to Taq polymerase, in alleles sequenced and cloned from two regions of *APC* exon 15 from lbl-1261, lbl-1260, lbl-c5 and lbl-a. Frequency of mutations observed in individual alleles cloned from *APC* gene regions (observed errors) are compared to the frequency expected due to Taq polymerase error in the PCR step (expected errors).

APC **Region**	**Cell line**	**Expected errors**	**Observed errors**	**Significance**	
B1-D2	lbl-1261	6.7	13	$\chi 2=1.8$,	p=0.18
	lbl-1260	6.5	54	$\chi 2=36.23$	p<0.001
	lbl-a	5.2	39	$\chi 2=26.3$	p<0.001
	lbl-c5	3.5	24	$\chi 2=14.21$	p<0.001
G1-I2	lbl-1261	7.1	11	$\chi 2=0.89$	p=0.346
	lbl-a	2.9	20	$\chi 2=12.56$	p<0.001

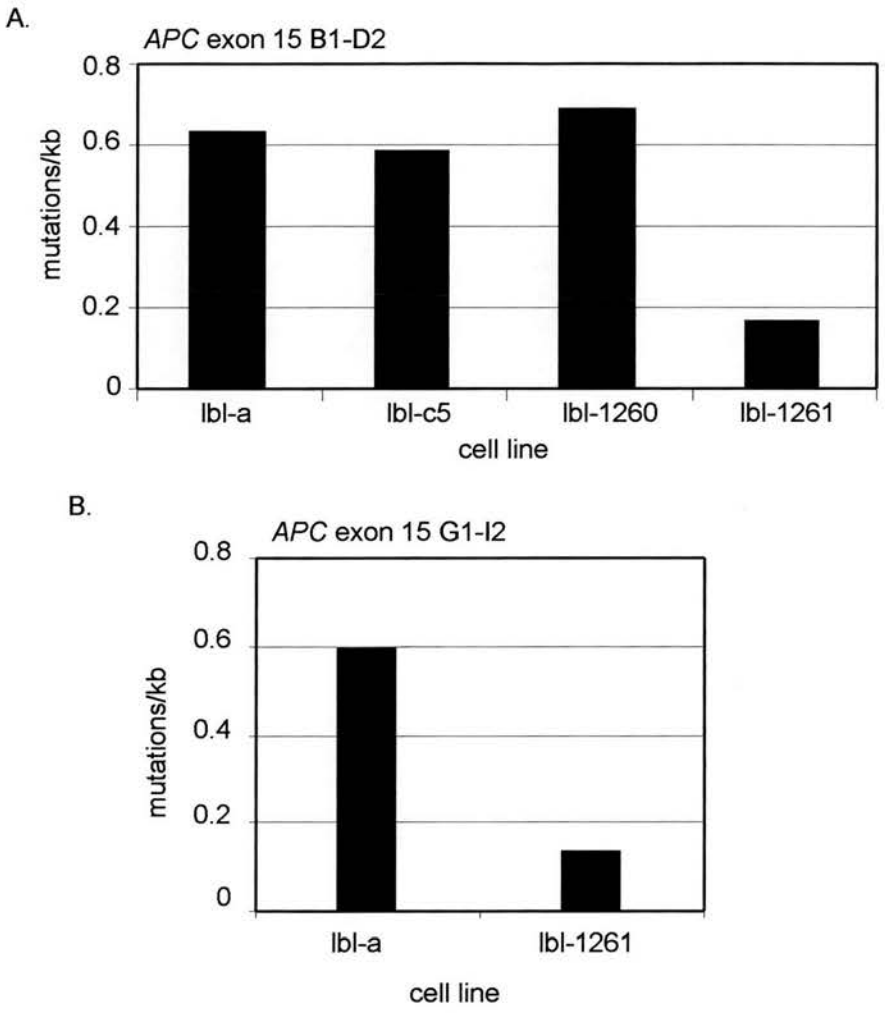


Figure 8.3 Relative mutation frequency detected in individual alleles cloned from lbl-1261, lbl-1260, lbl-c5 and lbl-a at two regions of the APC gene. (A) Mutation frequency over 910bp encompassed by region B1-D2 of *APC*. (B) Mutation frequency over 1114bp encompased by region G1-I2. At both regions the level of mutation is significantly elevated in MMR proficient cell lines, lbl-c5 and lbl-a compared to MMR deficient cell line, lbl-1261 ($p<0.001$).

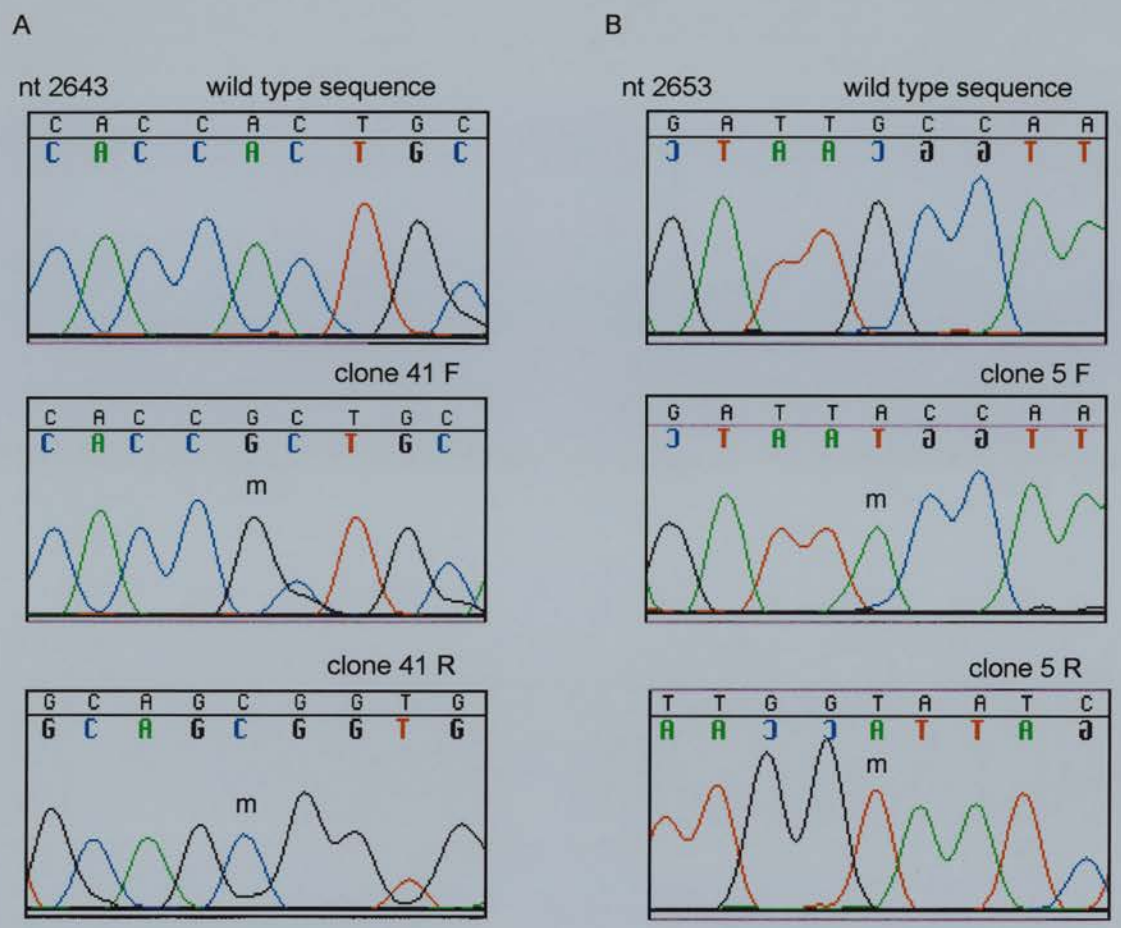


Figure 8.4 Representative mutations (m) in two alleles cloned from lbl-c5 template DNA at *APC* exon 15, region B1-D2. (A) An A to G transition was identified at nt2647 in clone 41. The top panel shows the wild type sequence at this site. The subsequent panels show the mutation revealed in both the forward and reverse sequence. (B) A transition of G to A was identified at nt2657 in clone 5. Again the top panel illustrates the wild type sequence at this site. Subsequent panels show that the mutation was confirmed in both directions.

8.3.4 Mutational spectrum at the *APC* gene in MMR proficient and MMR deficient lymphoblast cell lines

It was hypothesised that if the differences in mutation levels observed between lbl-1261 and MMR proficient cell lines were a result of different mechanisms of mutation, then the spectrum of mutations may also show significant differences. A propensity of frameshift mutations in lbl-1261 may be indicative that they arose as a consequence of MMR defects. Additionally if the inherent instability observed in the MMR proficient cell lines and lbl-1260 were due to similar factors, then the spectrum of mutations may also be expected to show similarities.

A detailed and extensive comparison of the nature and location of all 162 *APC* mutations in the lymphoblastoid cell lines analysed, is presented in Appendix A and Figure 8.5. Forward and reverse sequencing overlapped 41 of the mutations, confirming their presence in both directions. No false positives or false negatives were identified in any of the sequencing.

Huang *et al* previously reported differences in the mutational spectrum at the *APC* gene between MSI^+ and MSS tumours (Huang *et al.,* 1996). In contrast to the study by Huang *et al* there was no observable propensity for frameshift mutations within either of the two MMR deficient cell lines and no observable tendency for mutations to occur within simple repeated sequences in any of the cell lines analysed (Appendix A and Figure 8.5). Except for a single transversion mutation in cell lines lbl-1261 and lbl-a, all the identified mutations were transitions. However, Huang *et al* also reported a predominance of C→T changes in transition mutations in MSI^+ tumours (Huang *et al.,* 1996). It is of interest that in MMR deficient cell line lbl-1261 the frequency of both C→T and G→A transitions was significantly higher than in the MMR proficient control cell lines, lbl-a and lbl-c5 at both B1-D2 and G1-I2 ($p<0.0015$ at both regions) (Appendix A and Figure 8.5). Indeed the propensity for T→C and A→G transitions in lbl-a and lbl-c5 was not observed in lbl-1261 and this was seen most dramatically at region G1-I2 (Figure 8.5). Therefore, the significant difference in mutation frequency between lbl-1261 and MMR proficient control cell lines appears to be associated with a significant difference in the frequency of C→T

and G→A transition mutations. This may indicate underlying differences in the mechanisms resulting in mutation at the *APC* gene in lbl-1261 and MMR proficient lymphoblastoid cell lines. The similar frequency of *APC* mutations between lbl-1260 and MMR proficient control cell lines is also associated with a similar mutation spectrum, with lbl-1260 sharing the predominance for T→C and A→G mutations observed in the MMR proficient cell lines.

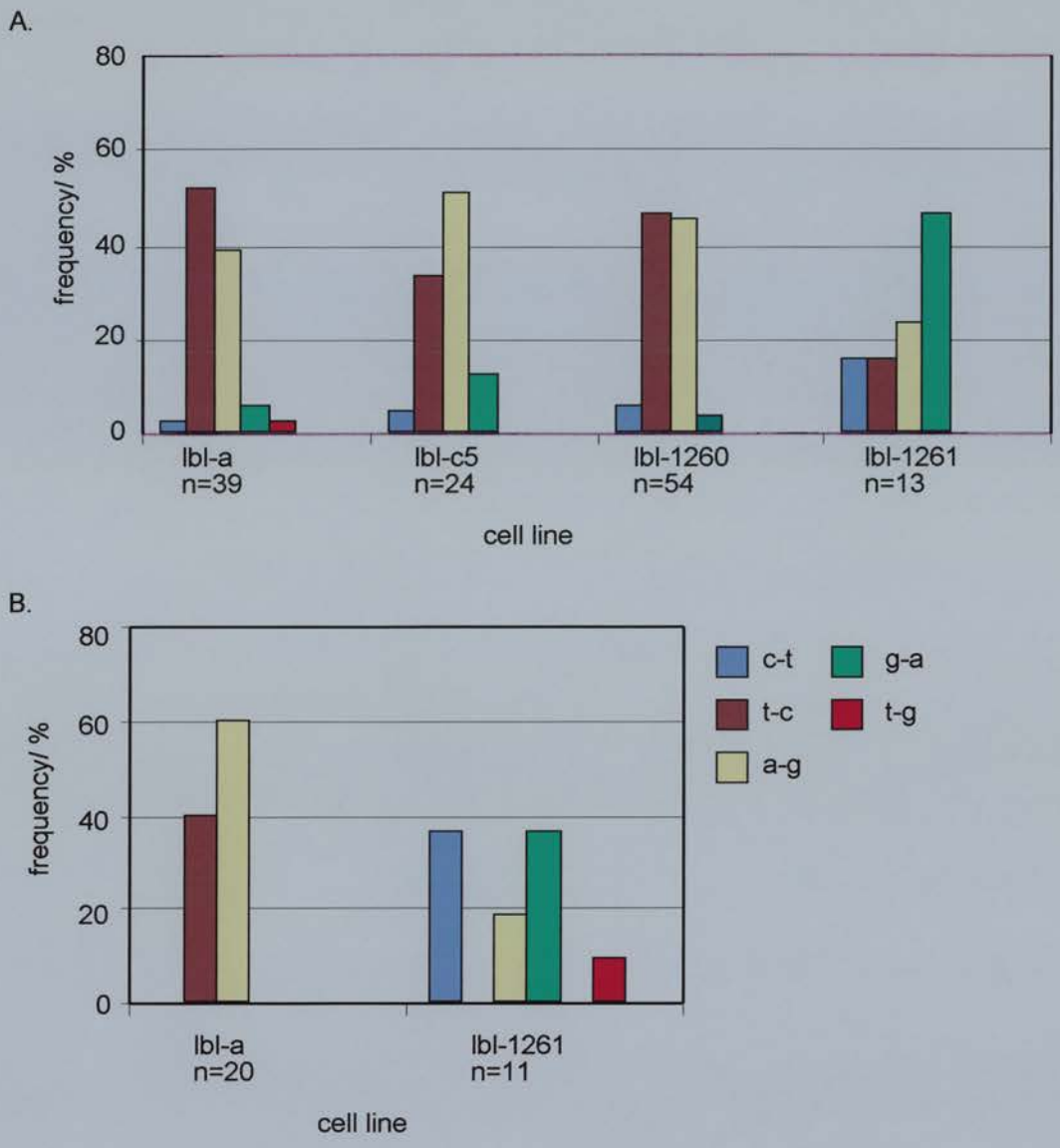


Figure 8.5 Spectrum of mutations identified in individual alleles cloned from two regions (B1-D2, G1-I2) of exon 15 of the *APC* gene in five lymphoblast cell lines. lbl-a and lbl-c5 are MMR proficient lymphoblastoid cell lines and lbl-1260 and lbl-1261 are MMR deficient lymphoblastoid cell lines. The frequency of each mutation is given in each cell line as a proportion of the total number (n) of mutations identified in that cell line.

8.3.5 Inherent instability at exon 15 of the *APC* gene in MMR proficient and MMR deficient cancer cell lines.

The initial data in 8.3.3 suggests that the *APC* gene is inherently prone to mutation but that in a highly unstable MMR deficient cell line instability at the *APC* gene is not detected. To further investigate this notion, inherent instability at the *APC* gene was examined in cancer cell lines with and without MMR defects. MMR proficient cell line SW480 (Parsons *et al.*, 1993), was analysed for inherent instability at region B1-D2 of the *APC* gene to address whether mutations were detected with a similarly significantly high frequency, as observed in MMR proficient lymphoblastoid cell lines lbl-a and lbl-c5. In addition, regions B1-D2 and G1-I2 were also analysed in the cancer cells HCT116 and HCT116+chromosome3 (HCT116+chr3). HCT116 is deficient for wild type MLH1 (Figure 3.2), exhibits MSI and is found to be defective in MMR (Parsons *et al.*, 1993; Koi *et al.*, 1994). However, the cell line HCT116+chr3 has restored MMR activity by the re-introduction of the *MLH1* gene on chromosome 3 (Koi *et al.*, 1994). These cell lines were used to provide a direct comparison of the effect from presence or absence of a functional MLH1 protein on the predilection for mutation at the *APC* gene.

147 alleles (120,319bp) cloned from SW480 template DNA, were sequenced and analysed at region B1-D2 (Table 8.6). A total of 79 mutations were identified, a mutation frequency of 0.66 mutations/kb (Table 8.6 and Figure 8.6). This frequency was very similar to that observed for MMR proficient lymphoblastoid cell lines and significantly greater than that observed in MMR deficient cell line lbl-1261 ($\chi 2=25.52$, $p<0.001$) (Figure 8.6 and Table 8.4). These data indicate that a MMR proficient cancer cell line also exhibits instability at the *APC* gene at a level significantly above that observed in a MMR deficient cell line derived from normal tissue. A comparison of mutation frequency between all MMR proficient cell lines analysed at the *APC* gene and other coding regions examined by identical methodology in this thesis was made (Table 8.7). This revealed a statistically significant difference between the frequency of mutation between the *APC* gene and other regions analysed ($\chi 2=13.23$, $p<0.001$). The statistically significant difference

in mutation frequency between MMR proficient cell lines at different loci strongly suggests that the majority of mutations detected at the *APC* gene are genuine and cannot be accounted for by explanations such as polymerase error.

At region B1-D2, a mutation frequency of 0.23 mutations/kb was determined for MMR deficient cancer cell line HCT116. However, analysis of cell line HCT116+chr3 did not reveal a significant increase in the number of mutations compared to HCT116 at region B1-D2 despite having an intact *MLH1* allele ($\chi 2=0.214$, $p=0.644$) (Table 8.6 and Figure 8.6). Furthermore, there was no significant difference in the level of mutations between these cell lines at the second *APC* region analysed (G1-I2) ($\chi 2=1.450$, $p=0.229$) (Table 8.6, Figure 8.6). The number of mutations at B1-D2 compared to G1-I2 also differed significantly ($p<0.001$, HCT116, $p=0.031$, HCT116+chr3). At region G1-I2, HCT116 and HCT116+chr3 had an elevated frequency of mutations compared to that at region B1-D2 ($\chi 2=0.473$, $p=0.491$) (Table 8.6). Therefore, although striking differences in inherent instability at the *APC* gene between HCT116 and HCT116+chr3 were predicted from the initial experiments in lymphoblastoid cell lines, this was not observed in the subsequent analysis.

Table 8.6 Analysis of individual alleles cloned and sequenced from two regions of exon 15 of the *APC* gene in cancer cell lines SW480, HCT116 and HCT116+chr3 (B1-D2, 910bp and G1-I2, 1114 bp). HCT116 is MMR deficient and SW480 and HCT116+chr3 are MMR proficient.

Region	Cell line	No. clones	No. Full length clones	Missing seq./kb	Total seq./ kb	No. mutations	Mutation freq. (mutations /kb)
B1-D2	SW480	147	103	13.451	120.319	79	0.66
	HCT116	96	77	1.836	85.497	20	0.23
	HCT116 +chr3	93	65	73	77.330	21	0.27
G1-I2	HCT116	70	47	11.768	66.212	42	0.63
	HCT116 +ch3	82	59	12.973	78.375	38	0.48

Table 8.7 Comparison of levels of inherent instability in MMR proficient cell lines at coding sequence loci in the *APC*, *CTNNB1* and *TGFBR2* genes. HCT116+chr3 is not included since it is MMR proficient only due to a chromosome transfer. *CTNNB1* ex3 refers to the region detailed in 8.3.1. *TGFBR2* ex3 refers to the non-repetitive region of exon 3 in *TGFBR2* as detailed in Chapter 7.

Locus	MMR proficient cell line	Mutation frequency (mutations/kb)	***APC*** **vs non *APC***
APC B1-D2	lbl-a	0.63	$\chi 2=13.23$, $p<0.001$
	lbl-c5	0.58	
	SW480	0.66	
APC G1-I2	lbl-a	0.59	
CTNNB1 ex3	lbl-a	0.108	
TGFBR2 ex 3	lbl-c5	Undetectable	

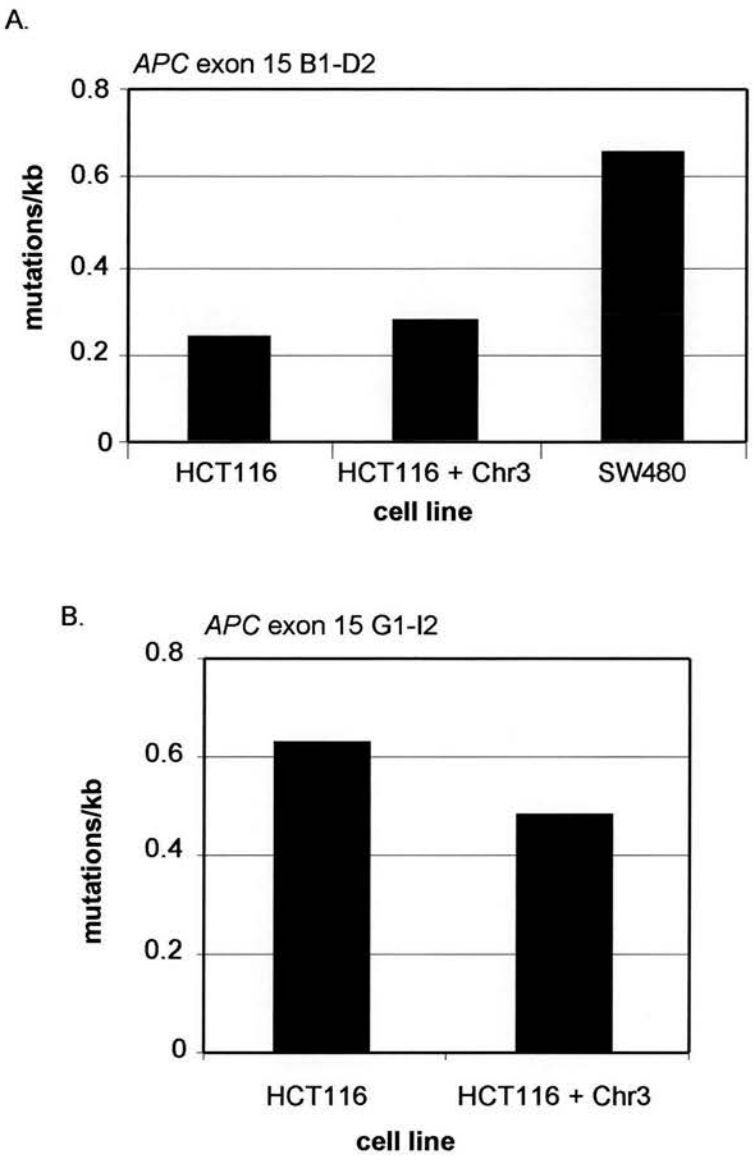


Figure 8.6 Relative mutation frequency detected in individual alleles cloned from cancer cell lines HCT116, HCT116+chromosome 3 and SW480 at two regions of the *APC* gene. (A) Mutation frequency over 910bp encompassed by region B1-D2. (B) Mutation frequency over 1114bp encompassed by region G1-I2.

8.3.6 Mutational spectrum at the *APC* gene in cancer cell lines

Overall, the 79 mutations detected in region B1-D2 of SW480 displayed a similar spectrum to those observed in MMR proficient lymphoblastoid cell lines lbl-a and lbl-c5 (Figure 8.7 and Appendix A). Again transition mutations predominated showing a propensity for T→C and A→G changes. A single transversion mutation was detected in SW480 at region B1-D2.

In HCT116 and HCT116+chr3 a total 121 mutations were identified. Again the majority were transition mutations in both HCT116 (59/62, 95%) and HCT116+chr3 (56/59, 95%) (Figure 8.7, Appendix A). Two transversions and a deletion were identified in region G1-I2 of HCT116, a single insertion was identified in region B1-D2 of cell line HCT16+chr3 and an insertion and a deletion at G1-I2. Although MMR deficient lymphoblast cell line lbl-1261 showed a significant difference in the frequency of C→T and G→A transitions compared to MMR proficient cell lines, such a difference between HCT116 and HCT116+chr3 was not observed ($p>0.1$ in all tests).

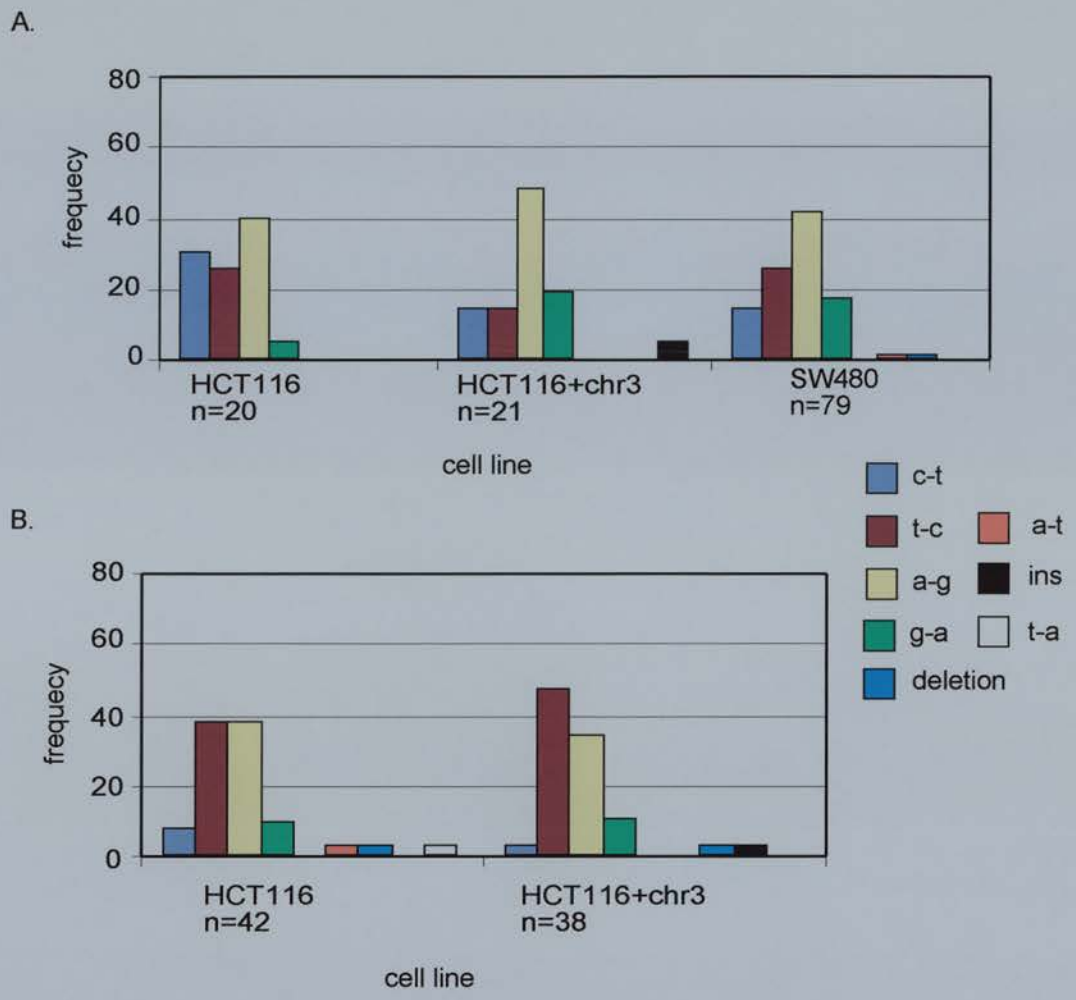


Figure 8.7 Spectrum of mutations in individual alleles at exon 15 of the APC gene in cancer cell lines HCT116, HCT116+chromosome 3 and SW480. (A) Region B1-D2. (B) Region G1-I2. The frequency of each mutation is given in each cell line as a proportion of the total number (n) of mutations identified.

8.4 Discussion

The investigations presented in this chapter set out to address whether mutations arise in specific coding sequences in β-catenin/TCF pathway as a consequence of DNA MMR defects. In this way it may be possible to infer the role of the β-catenin/TCF pathway in the progression of MSI^+ CRCs.

The regulatory domain of the *CTNNB1* gene and exon 15 of the *APC* gene have been reported as being mutated in a proportion of MSI^+ CRCs (Huang *et al.*, 1996; Homfray *et al.*, 1998; Sparks *et al.*, 1998; Mirabelli-Primdahl *et al.*, 1999; Miyaki *et al.*, 1999). However, in this analysis, there is no evidence to suggest that these regions are inherently prone to mutation exclusively as a result of MMR defects. Nonetheless, several intriguing observations have been made that may provide considerable insight into the susceptibility to mutation of sequences that are mutated in both MSI^+ and MSS colorectal cancers.

Analysis of individual alleles in exon 3 of the *CTNNB1* gene did not reveal an excess of mutations in a cell line derived from normal tissue with MMR defects (lbl-1261). The data suggest that other factors consequent of the process of tumourigenesis contribute to the observed accumulation of the mutations within this region in some MSI^+ CRCs (Akiyama *et al.*, 1997; Kitaeva *et al.*, 1997; Mirabelli-Primdahl *et al.*, 1999; Miyaki *et al.*, 1999). It should be noted that although the complete deletion of exon 3 has been observed in some MSI^+ CRCs (Iwao *et al.*, 1998) these lesions would not be detected by the technique used here.

Excess mutations were not observed at a low level in the $(GT)_4$ repeat of exon 10 of the *APC* gene. This region was previously shown to exhibit mutational mosaicism in the parent of a patient with FAP (Farrington and Dunlop, 1999). It was suggested that this repeat may be susceptible to mutation and the mosaicism might be associated with genetic instability. However, there was no evidence that mutations accumulate at this region consequent of MMR defects in this study. The previous study did not find constitutional defects in the MMR genes *MLH1* or *MSH2,* although it is possible that mutation in one of the other MMR genes contributed to the observed mosaicism. In addition, this *APC* repeat has not been reported as being

particularly prone to mutation in either MSI^+ or MSS cancers (Laurent-Puig *et al.*, 1998).

The observation that two MMR proficient (lbl-c5, lbl-a) lymphoblastoid cell lines, and a MMR proficient cancer cell line (SW480) displayed a significantly greater level of mutation at two regions of exon 15 of the *APC* gene compared to MMR defective cell line, lbl-1261 is intriguing. Furthermore, the level of mutation in MMR proficient control cell lines is significantly greater than that observed at other coding sequence loci such as exon 3 of the *TGFBR2* gene and exon 3 of the *CTNNB1* gene (Table 8.7). The data suggests that a region of the *APC* gene that is frequently mutated in around 80% of CRCs (Miyoshi *et al.*, 1992; Powell *et al.*, 1992), is inherently prone to mutations, even in cells that are derived from normal tissue. This would indicate that low but appreciable levels of mutations occur at these regions but that these are not corrected by MMR. Clearly, further validation of this result is merited but it is tempting to speculate that inherent susceptibility to mutation across the *APC* gene contributes to the frequent mutations found early during MSS tumourigenesis (Miyoshi *et al.*, 1992; Powell *et al.*, 1992; Miyaki *et al.*, 1994). The mutations observed were mainly point mutations and these lesions are commonly observed in CRCs (Miyoshi *et al.*, 1992; Powell *et al.*, 1992; Miyaki *et al.*, 1994; Huang *et al.*, 1996; Laurent-Puig *et al.*, 1998). However, whereas C→T transitions are the most often detected somatic *APC* point mutation in tumours (Laurent-Puig *et al.*, 1998), T→C and A→G transitions predominated here. A detailed analysis of the mutation spectrum between MMR proficient and MMR deficient cell lines that are non-cancer derived, did not identify any differences in the propensity for frameshift mutations or for mutation at SSRs within the *APC* gene as was reported previously (Huang *et al.,* 1996). However, in MMR deficient cell line lbl-1261, the frequency of C→T transitions differed significantly from MMR proficient cell lines, a similar observation to that reported by Huang *et al.,* (Huang *et al.,* 1996). This may suggest that the MMR defect in lbl-1261 modifies the mutation spectrum observed in this cell line.

The statistically significant increase in *APC* mutations in MMR proficient cell lines compared to lbl-1261 indicates that the majority of the mutations detected are genuine. Furthermore the significant differences in mutation frequency in MMR

proficient cell lines at alternative coding loci are difficult to reconcile with explanations such as error introduced in the PCR step of the cloning protocol. Indeed error attributable to Taq polymerase during PCR cannot account for the high frequency of *APC* mutations observed in the MMR proficient cell lines (Table 8.5). Repeated sequencing and re-isolation of plasmid stocks confirmed the identified mutations in every case. No instances of false positive or false negative sequencing errors were detected, suggesting sequencing error was not a major factor.

Statistical significant heterogeneity in mutation frequency has been demonstrated between lbl-1260 and lbl-1261 at microsatellite markers BAT-40 and D2S123 and also at the coding poly(A)$_{10}$ repeat of the *TGFBR2* gene in experiments presented in earlier chapters (Chapters 4 and 7). Further support for mutational heterogeneity between lbl-1260 and lbl-1261 dependent upon the nature of the inherent MMR defect is provided by the analysis of the *APC* gene at region B1-D2. At the poly(A)$_{10}$ repeat in *TGFBR2* lbl-1260 displayed a mutation frequency similar to that of control MMR proficient lymphoblast cell lines. This is also seen in the analysis here at the *APC* gene although it remains intriguing that the mutation frequency in lbl-1260 and MMR proficient cell lines is greater than in lbl-1261 whereas at other loci analysed it was lower.

Underlying differences in instability in lbl-1260 and lbl-1261 may contribute to the differences in frequency of mutation at *APC*. Indeed influence from mutation threshold may go some way to explaining the different frequency of mutations observed in the MMR deficient cell lines lbl-1261 and lbl-1260 at region B1-D2 of the APC gene. It has been proposed that the level of instability is crucial in determining whether a cell will be provided with a selective advantage (Cahill *et al.*, 1999). It is envisaged that if the accumulated damage from cell division rises above the threshold for viability, apoptotic pathways are activated and cell death ensues (Cahill *et al.*, 1999). Profound MMR defects in lbl-1261 as evidenced from the analysis of microsatellite loci (Chapter 4) may lead to increased mutation accumulation at the *APC* gene that is already inherently unstable. This combined with mutations at other coding regions (such as *TGFBR2*) may raise the level of mutation above the threshold for cell viability leading to apoptosis, with the remit being selection against any further *APC* mutations. In lbl-1260, the lower level of

instability observed at microsatellite sequences (Chapter 4), indicates a less profound defect in MMR. This may result in an increased level of mutation at *APC* but not to a level that exceeds the threshold for cell viability. It is of interest to note that of all the cell lines analysed, lbl-1260 had the highest frequency of mutation, although the mutational spectrum did not necessarily indicate that they were of the type associated with MMR defects.

It is also possible that resistance to DNA damage in the case of lbl-1260 may play an additional factor. Previous studies have suggested that the mismatch repair proteins, particularly MLH1 and MSH2, may be components of a pathway that influences apoptosis, since loss of apoptosis results from MLH1 or MSH2 deficiency (Zhang *et al.*, 1999). In addition, numerous studies have demonstrated that MLH1 deficient tumours are resistant to the DNA damaging agent MNNG and that this resistance is conferred via apoptotic pathways (Fink *et al.*, 1998; Duckett *et al.*, 1999; Hardman *et al.*, 2001). Since lbl-1260 has a defect in the *MLH1* gene, resistance to DNA damage may allow the accumulation of excess mutations but fail to induce apoptotic pathways. These may also be a factor in the analysis of cancer cell line HCT116, which also has a defect in *MLH1*. In contrast, there is evidence that apoptotic defects are not associated with defects in the *PMS2* gene and thus apoptosis in response to DNA damage in lbl-1261 may remain a viable mechanism (Zhang *et al.*, 1999). Therefore, it is speculated that a lower level of underlying genetic instability and/or resistance to apoptosis may contribute to the increased mutation frequency observed at the *APC* gene in lbl-1260 compared to lbl-1261.

The data from the two cancer cell lines HCT116 and HCT116+chr3 add further complexity to the observations presented above. Despite the fact that HCT116+chr3 is proficient in MMR, it did not display a significantly increased level of mutation compared to the same cell line in which MMR is deficient (HCT116). Furthermore, the lack of consistency in the results from each cell line at the two *APC* regions studied makes it difficult to draw conclusions from these results. These data illustrate the limitations of using cancer cells to address the direct consequences of MMR and point back to the conflicting reports of mutation at the *APC* gene in MSI^+ CRCs (Huang *et al.*, 1996; Homfray *et al.*, 1998). These data further justify the use of MMR deficient cells derived from normal tissue to unmask inherent instability

dependent on MMR defects. Another confounding factor is that although HCT116 is MMR deficient due to MLH1 defects, it also has other abnormalities. Mutation in the poly(C) tract of the *MSH6* gene has been identified, and the *MSH3* gene is also known to be mutated (Farrington, pers. comm.). Additional variabilities may contribute to the mutation frequency at any given locus making the dissection of events resulting exclusively from the MLH1 defect difficult. In addition, inhibition of wild type MLH1 expression in HCT116+chr3 using antisense RNA does not impair MMR activity (Chauhan *et al.*, 2000). This is suggestive of the presence of other unidentified factors that may contribute to the results presented here.

The notion that the *APC* gene may be inherently unstable even when MMR is intact clearly needs to be further clarified. Techniques such as single cell PCR may provide a complementary method to confirm the inherent instability observed in these cell lines. In addition, analysis of normal colon mucosa from MSS cancer patients, and also analysis of normal tissue from MMR deficient and MMR proficient mice, may also provide evidence that the *APC* gene is inherently susceptible to mutation in primary tissue.

In this chapter an alternative approach has been used to evaluate the role of mutations in β-catenin/TCF pathway genes in MMR deficient tumourigenesis. By utilising MMR deficient and MMR proficient cells derived from normal tissue, the inherent level of instability at three regions of the *APC* gene and exon 3 of the *CTNNB1* has been unmasked. The results suggest that the *APC* gene may be inherently susceptible to mutation at a low but appreciable level in cells that are proficient in MMR. This may reflect the high level of somatic mutation in the *APC* gene, observed in MSS colorectal cancers (Miyoshi *et al.*, 1992; Powell *et al.*, 1992) and the high new mutation rate for FAP (Bisgaard *et al.,* 1994). In the presence of MMR defects, underlying influences from instability may then modify the frequency with which mutations accumulate at the *APC* gene. Factors such as a mutational load within the cell and resistance to apoptosis may contribute to the propensity for further mutations to accumulate. The contribution of such factors would be dependent on the nature of the underlying MMR defects and may explain the conflicting frequencies of *APC* mutations identified in MSI^+ tumours. The relationship between MMR deficiency and susceptibility to mutation at the *APC*

locus appears complex. It will be of considerable interest to address the hypotheses that have been raised as a consequence of these data.

Chapter 9

Discussion and Summary

9.1 Summary

The work presented in this thesis has focused on molecular events that accumulate as a result of defects in MMR. Factors that influence the manifestation of MSI but which are not dependent on the process of tumourigenesis have also been determined.

The use of cell lines lbl-1260 and lbl-1261 has facilitated this research. In Chapter 3 these cell lines were demonstrated to be derived from B-cell lymphocytes and abnormal MMR protein expression, consistent with the profound MMR defects these cell lines are known to harbour, has been shown. The MMR defects in lbl-1260 and lbl-1261 are demonstrated to be associated with low level MSI even though they are not cancer derived (Chapter 4). This research has provided significant understanding of the analytical characteristics of two microsatellite markers used in the assessment of MSI status in CRCs (D2S123 and BAT-40). Examination of inherent instability of BAT-40 and D2S123 in lbl-1260 and lbl-1261 has revealed that the MSI phenotype exhibits significant heterogeneity. In particular, sequence susceptibility to mutation is influenced by intra-allelic differences that are dependent on constitutional phenotype. Allelic specific interruptions within the (CA)n repeat of microsatellite D2S123, confer stability to mutation in the presence of MMR defects even in MSI^+ tumours (Chapter 5). In addition, extreme susceptibility to mutation of the poly(A/T) repeat in BAT-40 has been revealed, by the demonstration that it is also inherently hypermutable in the germline (Chapter 6).

Investigation into the occurrence of mutations that are usually corrected by MMR at coding sequences in *TGFBR2, BAX, APC* and *CTNNB1* has provided considerable insight into the contribution that inherent sequence instability makes, to the accumulation of mutations in genes implicated in the development of MSI^+ CRCs. Differences in the mutation frequency of the repetitive tracts in *TGFBR2* and *BAX* in

MSI^+ CRCs has been demonstrated to reflect differences in the predilection of these repeats to instability when MMR is defective but when other effects of tumourigenesis are minimised. The $poly(A)_{10}$ repeat of *TGFBR2* has been specifically shown to be frequently mutated consequent of MMR defects in cells that are derived from normal tissue. β-catenin/TCF pathway genes *CTNNB1* and *APC* are mutated frequently in MSS CRCs and in some MSI^+ CRCs. Extensive analysis of coding sequence within these genes has not found any evidence that they are inherently prone to mutations as a direct result of MMR defects. However, the relationship between inherent stability at the *APC* locus and MMR deficiency appears to be complex. MMR defects may result in modification of the frequency and spectrum of mutations at the *APC* gene which itself is suggested to be inherently unstable even when MMR is proficient.

9.2 Study of molecular consequences of MMR defects using cell lines derived from normal tissue: an effective assay system.

Few studies have evaluated the relative contribution of inherent stability and selection, to the accumulation of mutations that are detected frequently in MMR deficient CRCs. The research in this thesis has utilised a sensitised in-vivo system, which effectively minimises selection and this has allowed inherent sequence stability to be unmasked. The system has facilitated the detection of genetic events that are specifically associated with defects in MMR, but that are not dependent on other influences associated with tumourigenesis itself.

The experimental rationale, with respect to the use of cell lines lbl-1260 and lbl-1261 has been validated during the course of the research. The use of these cell lines that are derived from normal tissue, but are defective in MMR, has effectively revealed molecular events that specifically accumulate, consequent of such defects. This approach has provided an alternative to the analysis of MSI^+ cancer cells. lbl-1260 and lbl-1261 represent a model system and as such the effects from selection cannot be entirely negated. However, the very fact that these cells are derived from non-cancerous tissue means that mutational bias from selection will be much reduced

compared to that in cancer cells. Furthermore, confounding molecular abnormalities consequent of the tumourigenic process itself will be largely eliminated. Striking differences in mutation frequency were detected between control (MMR proficient) cell lines and lbl-1260 and lbl-1261, at microsatellite repeats (D2S123 and BAT-40) and coding sequence (*TGFBR2* and *APC*). Since all the cell lines used were subject to the same conditions in culture, the data are not easily reconciled with explanations relating to selection or culture effects and this point was specifically discussed in Chapter 7. Furthermore, the stringent experimental controls and rigorous methodological techniques used throughout the experiments, means that technical error is likely at most, to be a minor contributor to the results presented. The most significant biological difference between the cell lines analysed, is the presence, absence or nature of MMR defects. The observed mutational differences are most easily explained as being consequent of these cell line specific differences in MMR status.

Previous reports have demonstrated MSI in normal tissues from patients with constitutional MMR defects (Miyaki *et al.*, 1997; Vilkki *et al.*, 2001; Wang *et al.*, 1999). However, these studies primarily focus on defining the clinical phenotype of such patients at a molecular level. The research in this thesis goes further, exploiting the value of these non-cancer derived MSI^+ cells by utilising them as an in vivo assay tool. Further to the characterisation of lbl-1260 and lbl-1261, these cell lines were employed to address specific questions that were defined in the research aims (Chapter 1), relating to inherent instability of sequences mutated frequently in MMR deficient CRCs. These cell lines have provided an excellent model with which to understand the molecular consequences of MMR defects.

Cell lines such as lbl-1260 and lbl-1261, have potential to be utilised for broader investigation of inherent sequence instability. It will be of considerable interest to carry out similar studies to those presented here at other loci to gain a greater insight into the susceptibility of different sequences to mutation in the presence of MMR defects. Furthermore, these cell lines could be employed to define the rate of mutation at inherently unstable sequences, using techniques such as single cell cloning or single cell PCR.

9.3 Heterogeneity in MMR defects results in heterogeneity of the mutator phenotype: insight into mechanisms of MMR

Both lbl-1260 and lbl-1261 have been demonstrated to lack MMR activity on a biochemical level (Parsons *et al.*, 1995a). However, the data presented throughout this thesis consistently demonstrates that these cell lines derived from normal tissue with different constitutional MMR defects have different mutator phenotypes in terms of the frequency and spectrum of mutations that arise.

While the dominant negative effect of the *PMS2* mutation in lbl-1261 has been demonstrated previously (Nicolaides *et al.*, 1998), the data in Chapter 3 provides compelling evidence that the defect in lbl-1260 is conferred by a reduction in MLH1 protein expression. Further study is necessary to establish the precise mechanism that prevents the translation of the *MLH1* transcripts in this cell line and to determine the contribution that the *MLH1* mutation in lbl-1260 makes to the observed reduction in protein expression. It will be of interest to establish whether expression of the wild type lbl-1260 *MLH1* allele at the protein level can be detected in the absence of the mutant. This could be determined, by cloning and transfecting the wild type allele into an MLH1 deficient cell line such as HCT116 and subsequent assay of nuclear extracts by western analysis.

Lbl-1261 displays dramatic instability at coding and non-coding repeat sequences, supporting previous evidence that shows MMR activity is completely abolished in this cell line (Parsons *et al.*, 1995a). Furthermore the detection of mutations at both mononucleotide (BAT-40, *TGFBR2*) and dinucleotide (D2S123) repeats is consistent with PMS2 functioning in the repair of base:base mismatches and insertion:deletion loops (Figure 1.3) (Li and Modrich, 1995; Kolodner and Marsischky, 1999).

It is well established that *MLH1* also functions in the repair of base:base mispairs and insertion:deletion loops (Kolodner and Marsischky, 1999) and the observation of mutation at BAT-40 and D2S123 in lbl-1260 is entirely consistent with this. However, the lower level of instability in lbl-1260 compared to lbl-1261 at both microsatellite repeat sequences and at the *TGFBR2* poly$(A)_{10}$ tract is intriguing. This may indicate that residual MMR activity is taking place, but that the expression of functional MLH1 is below the level of detection.

That the nature of the MMR defect contributes to heterogenity in the frequency and spectrum of accumulating mutations is not only evident from the analyses at non-coding and coding repeat sequences but also from the analysis at two regions of the *APC* gene. Significant differences in mutation frequency between lbl-1260 and lbl-1261 again suggest influence from differential repair capabilities in these two cell lines. However, at *APC* the mutation frequency appeared to be greater in cell line lbl-1260 than in lbl-1261 in contrast to the trend observed at other loci. Possible explanations for such a phenomenon were considered in Chapter 8.

Analysis of cells derived from normal tissue containing different MMR mutations to those in lbl-1260 and lbl-1261 will facilitate investigation into whether the nature of the MMR mutation or the gene that is mutated, is of greater importance in directing the frequency and spectrum of mutations that consequently accumulate. Such studies could also be carried out, utilising normal tissue from MMR deficient mice, especially since these tissue are relatively easy to acquire.

9.4 Susceptibility of microsatellite repeats to mutation is dependent on inherent instability and can be modified by loci dependent factors

The work presented in Chapters 4-6 provides fundamental understanding of the stability of non-coding microsatellite sequences in the human genome (Bacon *et al.*, 2000; Bacon *et al.*, 2001a; Bacon *et al.*, 2001b). Furthermore, since the loci investigated are used in CRC MSI screening strategies, the results have relevance to the effective implementation of such programs and to the correlation of cancer with clinical outcome.

The data supports the notion that the likelihood a microsatellite will be susceptible to instability may relate to the inherent mutation rate at that locus (Thibodeau *et al.*, 1998). The detection of mutations in lbl-1260 and lbl-1261 at BAT-40 and D2S123 that are specifically associated with MMR defects (Chapter, 4) indicates that these loci are inherently unstable by nature of the fact that these cell lines are not cancer-derived. Further evidence for the relationship between inherent sequence instability

and susceptibility to mutation in MMR defective CRCs has been presented in Chapters 5 and 6.

In Chapter 5, a TA interruption in the (CA)n repeat of D2S123 is demonstrated to confer stability on the allele. Constitutional genotype is shown to determine the inherent susceptibility of this locus to mutation and directly affects whether D2S123 will accumulate mutations in MMR deficient colorectal tumours (Bacon *et al.*, 2000). Due to the important clinical implications of this finding, it will be of considerable interest to re-asses studies in which tumour MSI status was determined using the panel of five recommended markers (Boland *et al.*, 1998). MSI-L or MSS colorectal tumours, from patients with inherently "stable" constitutional D2S123 genotypes, may have been inappropriately classified. This can be easily confirmed by analysis of tumour MSI at other microsatellite loci. Evaluating the extent of such mis-classification will only underscore the importance of assessing the analytical characteristics of microsatellite markers, prior to using them in assays for detecting tumour stability.

It is clear from the CEPH database that many microsatellite markers used in MSI analysis are polymorphic as determined by size analysis (www.cephb.fr/). Determination of allele specific sequence differences may highlight those loci at which constitutional genotype may influence susceptibility to mutation in MMR deficient tumours. Identifying intra-allelic repeat sequence differences could direct subsequent analysis of the inherent allele stability at a given locus, in cell lines such as lbl-1260 and lbl-1261 and in MMR deficient tumours. Such studies can only serve to increase the robustness of the MSI marker panel and the confidence with which these microsatellites are used.

This work has further relevance to the identification of functionally important coding sequences that may be susceptible to mutation in MMR deficient cancers. Repeat sequences are prone to mutation in coding as well as non-coding sequences in MSI^+ CRCs (Ionov *et al.*, 1993; Thibodeau *et al.*, 1993; Markowitz *et al.*, 1995; Parsons *et al.*, 1995b; Rampino *et al.*, 1997; Duval *et al.*, 2001). It is likely that further, functionally important coding sequence repeat loci will be identified as being frequently mutated in MSI^+ CRCs and implicated as contributing to cancer

development. Rigorous assessment of population polymorphism at such loci should be carried out. If allelic stability similar to that observed at D2S123 is conferred on such a coding sequence repeat, it could be envisaged that constitutional genotype may not only confer resistance to mutation consequent of MMR defects, but may also be functionally protective towards CRC development. It is also rational to assess the mutational stability of coding sequence loci logged in GenBank as short interspersed repeats e.g. (CA)nTA(CA)n or (A)nT(A)n. Allelic variants at such loci that lack an interspersion, may reveal longer repetitive tracts that are highly susceptible to mutation consequent of MMR defects and that are also functionally important.

The importance of inherent sequence stability on the likelihood of mutation accumulation in MMR deficient tumours is underpinned by the study in Chapter 6. BAT-40 is one of the most highly mutable microsatellite markers in MMR deficient CRCs (Parsons *et al.*, 1995b) and the work presented has provided evidence that this reflects extreme inherent instability at this locus. The profound susceptibility to mutation at BAT-40, identified in Chapter 4 is confirmed by the finding that this poly(A/T) repeat is highly unstable in the germline (Bacon *et al.*, 2001a). It will be of importance to further this founding observation, by defining the rate of mutation by more rigorous analysis such as single sperm cell PCR. In addition, further investigation as to whether this phenomenon is common to many poly(A/T) repeats, both in coding and non-coding sequence, will strengthen the initial finding made here and be both of considerable interest and fundamental importance.

BAT-26 is another poly(A/T) marker frequently employed in MSI analysis (Boland *et al.*, 1998). Despite being very sensitive to MSI in MMR deficient tumours, BAT-26 has been described as showing quasi-monomorphic variation in the population in contrast to the polymorphic nature of BAT-40 (Hoang *et al.*, 1997; Zhou *et al.*, 1997). It will be of interest to address whether BAT-26 displays germline instability as indicated by the high propensity for mutation in MMR deficient tumours, or whether the quasi-monomorphic population variance is correlated with stability in the germline (Hoang *et al.*, 1997; Zhou *et al.*, 1997).

It has already been discussed in Chapter 6 that germline instability of poly(A/T) loci may have functional consequences in view of the prevalence of poly(A/T) repeats in coding and non-coding sequence. Available sequence from the human genome was recently subjected to rigorous searching for coding sequence mononucleotide and dinucleotide tracts over 8 repeat units in length. The results of this search were filtered for redundancy and for the presence of pseudogenes. 215 coding sequence loci were identified containing poly(A/T) repeats of over 10bp, some with repeats over 30bp in length (Taylor pers. comm.). Many of these repeats were within putative genes whose function has yet to be established (Taylor pers. comm.). Thus if hypermutability of poly(A/T) sequences in the germline occurs frequently, there are many potential loci at which functional impact from this phenomenon could be conferred.

An in depth analysis of inherent stability at all the microsatellite markers that are used in the analysis of MSI status of CRCs is not trivial task (Boland *et al.*, 1998). However, the data presented in this thesis have demonstrated that such investigations are worthwhile and can significantly improve our understanding of how the mutator phenotype manifests in CRCs, as well as offering more reliable molecular tools for clinical and research purposes.

9.5 Inherent susceptibility of coding sequences to mutations contributes to the frequency with which they are mutated in MMR deficient CRCs.

The data presented in Chapters 7 and 8 has provided evidence that inherent stability of coding sequences, contributes to the frequency with which they are mutated in MMR deficient tumours and thus implicates a significant role for MMR deficiency in the development of CRC. The results are consistent with the hypothesis presented at the outset of this thesis that sequences mutated frequently in MMR deficient CRCs are inherently prone to mutation. The data strengthens support for the importance of an increased mutation rate in tumourigenesis (Loeb, 2001).

In particular compelling evidence is presented in Chapter 7 (Bacon *et al.*, 2001b). The data demonstrate that MMR deficiency results in an excess of mutations, specifically at the poly(A)$_{10}$ tract compared to other regions of the *TGFBR2* gene ($p<0.001$). Conversely, an excess of mutations does not appear to arise at the poly(G)$_8$ tract of the *BAX* gene. These studies provide insight into the mechanism by which *TGFBR2* and *BAX* become mutated during tumourigenesis. The work has suggested that differences in the mutation frequency observed at the repetitive tracts within *TGFBR2* and *BAX* in MSI$^+$ CRCs, are associated with differences in the susceptibility of these loci to mutation exclusively as a result of MMR defects. These findings support the idea that the relative contribution from selection and mutation may be different at alternative loci, which are mutated to appreciable levels in MSI$^+$ CRCs. *TGFBR2* in particular, is mutated with extreme consistency in MMR defective colorectal tumours and is implicated as having a critical role in early stages of tumourigenesis (Markowitz *et al.*, 1995; Parsons *et al.*, 1995b; Markowitz, 2000). The fact that the poly(A)$_{10}$ tract within *TGFBR2* is inherently susceptible to mutation, indicates that inherent mutation frequency as unmasked by MMR defects, plays a major role in the accumulation of these mutations in MSI$^+$ CRCs and thus contributes significantly to tumourigenesis (Bacon *et al.*, 2001b).

It will be of importance to establish whether other coding poly(A) repeats also accumulate mutations exclusively as a result of MMR defects, to a similar level as that observed in *TGFBR2*. This may indicate the relative importance of the predilection of the repeat in *TGFBR2* to mutation, in MSI$^+$ CRCs. Identification of other loci that are similarly prone to mutation consequent of MMR defects may reveal other functionally important loci that are abrogated in the progression of MSI$^+$ CRCs. As described above many coding sequence loci in the human genome contain poly(A) tracts over 10bp in length. These represent potential functional targets for mutation accumulation in MSI$^+$ CRCs. This work has shown that MMR defects specifically allow mutations to accumulate at one such coding poly(A) repeat. However, a systematic analysis of inherent stability of all coding sequence repeats and analysis of their frequency of mutation in MSI$^+$ CRCs would provide considerable insight into the role of MMR deficiency in tumourigenesis. Such analysis would further understanding of the relationship between inherent sequence

instability and the frequency of gene mutation in MSI^+ tumours. Although the total number of human coding repeat sequences, over 10 repeating units in length is appreciable, extensive mining of the human genome suggests the number does not pose an impossible research task (Taylor pers. comm.)

The examination of the susceptibility of two β-catenin/TCF pathway genes to mutation consequent of MMR defects, highlights that the relationship between the inherent instability of a sequence and the frequency with which it is mutated in MSI^+ CRCs is likely to be complex in many cases. Further investigation of the preliminary findings presented in Chapter 8 is clearly necessary. Nonetheless, this work adds to previous studies that have investigated the role that abrogation of the β-catenin/TCF pathway may play in the development of CRCs with MMR defects. The data presented does not find any evidence to suggest that mutations arise in the *APC* gene or the *CTNNB1* gene exclusively as a consequence of repair defects. Although, the initial research has resulted in the speculation that the *APC* gene is subject to a low but appreciable level of mutation in MMR proficient cells. This is a radical suggestion that clearly needs to be substantiated. However, the notion that *APC* may be readily prone to mutation even when MMR is functional, may explain why it is mutated frequently in sporadic CRCs (Miyoshi *et al.*, 1992; Powell *et al.*, 1992) and also the high new mutation rate in FAP patients (Bisgaard *et al.*, 1994; Farrington and Dunlop, 1999).

The work presented in this thesis provides insight into the relative contribution of inherent instability at sequences mutated frequently in MMR defective cancers and reveals that it is possible to dissect out those events that occur exclusively in the presence of MMR deficiency. This research offers promise for new levels of understanding of how MSI^+ cancers develop and the role that an increased mutation frequency may contribute to tumourigenesis. The revelation that loci specific and allele specific factors modify sequence susceptibility to mutation, significantly affecting the manifestation of the mutator phenotype, adds further complexity to the study of events consequent of MMR defects. However, identification of such factors at other loci will be critical in order to fully appreciate the contribution that inactivation of MMR makes to the development of cancer. Detailed understanding of the molecular consequences of MMR can only have a positive impact on

understanding how cancers associated with these defects develop and ultimately be of clinical and public health benefit.

Chapter 10 Bibliography

Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Powell, S.M., Jen, J., and Hamilton, S.R. (1993). Clues to the pathogenesis of familial colorectal cancer. Science, ***260***: 812-816.

Aaltonen, L.A., Peltomaki, P., Mecklin, J.P., Jarvinen, H., Jass, J.R., Green, J.S., Lynch, H.T., Watson, P., Tallqvist, G., and Juhola, M. (1994). Replication errors in benign and malignant tumors from hereditary nonpolyposis colorectal cancer patients. Cancer Res., ***54***: 1645-1648.

Aaltonen, L.A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomaki, P., Chadwick, R.B., Kaariainen, H., Eskelinen, M., Jarvinen, H., Mecklin, J.P., and de la Chapelle, A. (1998). Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. N. Engl. J. Med., ***338***: 1481-1487.

Abdel-Rahman, W.M., Georgiades, I.B., Curtis, L.J., Arends, M.J., and Wyllie, A.H. (1999). Role of BAX mutations in mismatch repair-deficient colorectal carcinogenesis. Oncogene, ***18***: 2139-2142.

Acharya, S., Wilson, T., Gradia, S., Kane, M.F., Guerrette, S., Marsischky, G.T., Kolodner, R., and Fishel, R. (1996). hMSH2 forms specific mispair-binding complexes with hMSH3 and hMSH6. Proc. Natl. Acad. Sci. U. S. A., ***93***: 13629-13634.

Akiyama, Y., Iwanaga, R., Saitoh, K., Shiba, K., Ushio, K., Ikeda, E., Iwama, T., Nomizu, T., and Yuasa, Y. (1997a). Transforming growth factor beta type II receptor gene mutations in adenomas from hereditary nonpolyposis colorectal cancer. Gastroenterology, ***112***: 33-39.

Akiyama, Y., Sato, H., Yamada, T., Nagasaki, H., Tsuchiya, A., Abe, R., and Yuasa, Y. (1997b). Germ-line mutation of the hMSH6/GTBP gene in an atypical hereditary nonpolyposis colorectal cancer kindred. Cancer Res., ***57***: 3920-3923.

Akiyama, Y., Nagasaki, H., Yagi, K.O., Nomizu, T., and Yuasa, Y. (2000) beta-Catenin and adenomatous polyposis coli (APC) mutations in adenomas from hereditary non-polyposis colorectal cancer patients. Cancer Lett., ***157***: 185-191.

Alani, E. (1996). The Saccharomyces cerevisiae Msh2 and Msh6 proteins form a complex that specifically binds to duplex oligonucleotides containing mismatched DNA base pairs. Mol. Cell Biol., ***16***: 5604-5615.

Alexandrow, M.G. and Moses, H.L. (1995). Transforming growth factor beta 1 inhibits mouse keratinocytes late in G1 independent of effects on gene transcription. Cancer Res., ***55***: 3928-3932.

Andreassen, R., Egeland, T., and Olaisen, B. (1996). Mutation rate in the hypervariable VNTR g3 (D7S22) is affected by allele length and a flanking DNA sequence polymorphism near the repeat array. Am. J. Hum. Genet., ***59***: 360-367.

Ashley, C.T.J. and Warren, S.T. (1995). Trinucleotide repeat expansion and human disease. Annu. Rev. Genet., ***29:*** 703-728.

Attisano, L. and Wrana, J.L. (1998). Mads and Smads in TGF beta signalling. Curr. Opin. Cell. Biol., ***10***: 188-194.

Au, K.G., Welsh, K., and Modrich, P. (1992). Initiation of methyl-directed mismatch repair. J. Biol .Chem., ***267***: 12142-12148.

Augenlicht, L.H., Richards, C., Corner, G., and Pretlow, T.P. (1996). Evidence for genomic instability in human colonic aberrant crypt foci. Oncogene., ***12***: 1767-1772.

Bacon, A.L., Farrington, S.M., and Dunlop, M.G. (2000). Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells. Hum. Mol. Genet., ***9***: 2707-2713.

Bacon, A.L., Dunlop, M.G., and Farrington, S.M. (2001a). Hypermutability of a poly(A/T) tract in the human germline. Nucleic Acids Res., ***29:*** 4405-4413.

Bacon, A.L., Farrington, S.M., and Dunlop, M.G. (2001b). Mutation frequency in coding and non-coding repeat sequences in mismatch repair deficient cells derived from normal human tissue. Oncogene, ***20:*** 7464-7471

Bader, S., Walker, M., Hendrich, B., Bird, A., Bird, C., Hooper, M., and Wyllie, A. (1999). Somatic frameshift mutations in the MBD4 gene of sporadic colon cancers with mismatch repair deficiency. Oncogene, ***18***: 8044-8047.

Baker, S.M., Bronner, C.E., Zhang, L., Plug, A.W., Robatzek, M., Warren, G., Elliott, E.A., Yu, J., Ashley, T., and Arnheim, N. (1995). Male mice defective in the DNA mismatch repair gene PMS2 exhibit abnormal chromosome synapsis in meiosis. Cell, ***82***: 309-319.

Ban, C. and Yang, W. (1998). Crystal structure and ATPase activity of MutL: implications for DNA repair and mutagenesis. Cell, ***95***: 541-552.

Bapat, B., Xia, L., Madlensky, L., Mitri, A., Tonin, P., Narod, S.A., and Gallinger, S. (1996). The genetic basis of Muir-Torre syndrome includes the hMLH1 locus. Am. J. Hum. Genet., ***59***: 736-739.

Barnetson, R., Jass, J., Tse, R., Eckstein, R., Robinson, B., and Schnitzler, M. (2000). Mutations associated with microsatellite unstable colorectal carcinomas exhibit widespread intratumoral heterogeneity. Genes Chromosomes Cancer, ***29***: 130-136.

Barth, A.I., Nathke, I.S., and Nelson, W.J. (1997). Cadherins, catenins and APC protein: interplay between cytoskeletal complexes and signaling pathways. Curr. Opin. Cell Biol., ***9***: 683-690.

Behrens, J., von Kries, J.P., Kuhl, M., Bruhn, L., Wedlich, D., Grosschedl, R., and Birchmeier, W. (1996). Functional interaction of beta-catenin with the transcription factor LEF-1. Nature, ***382***: 638-642.

Behrens, J. (2000). Control of beta-catenin signaling in tumor development. Ann. N. Y. Acad. Sci., ***910:*** 21-33.

Bellacosa, A., Cicchillitti, L., Schepis, F., Riccio, A., Yeung, A.T., Matsumoto, Y., Golemis, E.A., Genuardi, M., and Neri, G. (1999). MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proc. Natl. Acad. Sci. U. S. A., ***96***: 3969-3974.

Bellacosa, A., Genuardi, M., Anti, M., Viel, A., and Ponz, d.L. (1996). Hereditary non polyposis colorectal cancer: review of clinical, molecular genetics, and counselling aspects. Am. J. Med. Genet., ***62***: 353-364.

Bernstein, P. and Ross, J. (1989). Poly(A), poly(A) binding protein and the regulation of mRNA stability. Trends Biochem. Sci., ***14***: 373-377.

Bichara, M., Schumacher, S., and Fuchs, R.P. (1995). Genetic instability within monotonous runs of CpG sequences in Escherichia coli. Genetics, ***140***: 897-907.

Bird, A.P. (1986). CpG-rich islands and the function of DNA methylation. Nature., ***321***: 209-213.

Bisgaard, M.L., Fenger, K., Bulow, S., Niebuhr, E., and Mohr, J. (1994). Familial adenomatous polyposis (FAP): frequency, penetrance, and mutation rate. Hum. Mutat., ***3***: 121-125.

Blackwell, L.J., Martik, D., Bjornson, K.P., Bjornson, E.S., and Modrich, P. (1998). Nucleotide-promoted release of hMutSalpha from heteroduplex DNA is consistent with an ATP-dependent translocation mechanism. J. Biol. Chem., ***273***: 32055-32062.

Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N., and Srivastava, S. (1998). A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res., ***58***: 5248-5257.

Boland, C.R. (2000). Molecular genetics of hereditary nonpolyposis colorectal cancer. Ann. N. Y. Acad. Sci., ***910***: 50-59.

Bonelli, L., Martines, H., Conio, M., Bruzzi, P., and Aste, H. (1988). Family history of colorectal cancer as a risk factor for benign and malignant tumours of the large bowel. A case-control study. Int. J. Cancer., ***41***: 513-517.

Borresen, A.L., Lothe, R.A., Meling, G.I., Lystad, S., Morrison, P., Lipford, J., Kane, M.F., Rognum, T.O., and Kolodner, R.D. (1995). Somatic mutations in the hMSH2 gene in microsatellite unstable colorectal carcinomas. Hum. Mol. Genet., ***4***: 2065-2072.

Boyes, J. and Bird, A. (1992). Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J., ***11***: 327-333.

Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J., and Rolf, B. (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am. J. Hum. Genet., ***62***: 1408-1415.

Bronner, C.E., Baker, S.M., Morrison, P.T., Warren, G., Smith, L.G., Lescoe, M.K., Kane, M., Earabino, C., Lipford, J., and Lindblom, A. (1994). Mutation in the DNA mismatch repair gene homologue hMLH1 is associated with hereditary non-polyposis colon cancer. Nature, ***368***: 258-261.

Bull, L.N., Pabon-Pena, C.R., and Freimer, N.B. (1999). Compound microsatellite repeats: practical and theoretical features. Genome Res., ***9***: 830-838.

Burkitt, D.P. (1971). Epidemiology of cancer of the colon and rectum. Cancer, ***28***: 3-13.

Cahill, D.P., Kinzler, K.W., Vogelstein, B., and Lengauer, C. (1999). Genetic instability and darwinian selection in tumours. Trends Cell Biol., ***9***: M57-M60

Caldes, T., Perez-Segura, P., Tosar, A., De La Hoya, M., and Diaz-Rubio, E. (2000). Low frequency of microsatellite instability in sporadic breast cancer. Int. J. Oncol., ***16***: 1235-1242.

Carethers, J.M. and Pham, T.T. (2000). Mutations of transforming growth factor beta 1 type II receptor, BAX, and insulin-like growth factor II receptor genes in microsatellite unstable cell lines. In Vivo, ***14***: 13-20.

Chauhan, D.P., Yang, Q., Carethers, J.M., Marra, G., Chang, C.L., Chamberlain, S.M., and Boland, C.R. (2000). Antisense inhibition of hMLH1 is not sufficient for loss of DNA mismatch repair function in the HCT116+chromosome 3 cell line. Clin. Cancer Res., ***6***: 3827-3831.

Chen, J., Heerdt, B.G., and Augenlicht, L.H. (1995). Presence and instability of repetitive elements in sequences the altered expression of which characterizes risk for colonic cancer. Cancer Res., ***55***: 174-180.

Chillon, M., Casals, T., Mercier, B., Bassas, L., Lissens, W., Silber, S., Romey, M.C., Ruiz-Romero, J., Verlingue, C., and Claustres, M. (1995). Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. N. Engl. J. Med., ***332***: 1475-1480.

Chu, C.S., Trapnell, B.C., Curristin, S., Cutting, G.R., and Crystal, R.G. (1993). Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. Nat. Genet., ***3***: 151-156.

Chung, M.Y., Ranum, L.P., Duvick, L.A., Servadio, A., Zoghbi, H.Y., and Orr, H.T. (1993). Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat. Genet., ***5***: 254-258.

Chung, Y.J., Park, S.W., Song, J.M., Lee, K.Y., Seo, E.J., Choi, S.W., and Rhyu, M.G. (1997). Evidence of genetic progression in human gastric carcinomas with microsatellite instability. Oncogene, ***15***: 1719-1726.

Cohen, P.R., Kohn, S.R., and Kurzrock, R. (1991). Association of sebaceous gland tumors and internal malignancy: the Muir-Torre syndrome. Am. J. Med., ***90***: 606-613.

Crail, H.W. (1949). Multiple primary malignancies of arising in the rectum, brain, and thyroid. Nav. Med. Bull., ***49***: 123-127.

Crawford, D.C., Wilson, B., and Sherman, S.L. (2000). Factors involved in the initial mutation of the fragile X CGG repeat as determined by sperm small pool PCR. Hum. Mol. Genet., ***9***: 2909-2918.

Cross, S.H. and Bird, A.P. (1995). CpG islands and genes. Curr. Opin. Genet. Dev., ***5***: 309-314.

Cunningham, J.M., Christensen, E.R., Tester, D.J., Kim, C.Y., Roche, P.C., Burgart, L.J., and Thibodeau, S.N. (1998). Hypermethylation of the hMLH1 promoter in colon cancer with microsatellite instability. Cancer Res., ***58***: 3455-3460.

Cunningham, J.M., Kim, C.Y., Christensen, E.R., Tester, D.J., Parc, Y., Burgart, L.J., Halling, K.C., McDonnell, S.K., Schaid, D.J., Walsh, V.C., Kubly, V., Nelson, H., Michels, V.V., and Thibodeau, S.N. (2001). The frequency of hereditary defective mismatch repair in a prospective series of unselected colorectal carcinomas. Am. J. Hum. Genet., ***69***: 780-790.

De Rosa, M., Fasano, C., Panariello, L., Scarano, M.I., Belli, G., Iannelli, A., Ciciliano, F., and Izzo, P. (2000). Evidence for a recessive inheritance of Turcot's syndrome caused by compound heterozygous mutations within the PMS2 gene. Oncogene, ***19***: 1719-1723.

De Wind, N., Dekker, M., Berns, A., Radman, M., and te, R.H. (1995). Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. Cell, ***82***: 321-330.

Delattre, O., Olschwang, S., Law, D.J., Melot, T., Remvikos, Y., Salmon, R.J., Sastre, X., Validire, P., Feinberg, A.P., and Thomas, G. (1989). Multiple genetic alterations in distal and proximal colorectal cancer. Lancet, ***2***: 353-356.

Di Rienzo, A., Donnelly, P., Toomajian, C., Sisk, B., Hill, A., Petzl-Erler, M.L., Haines, G.K., and Barch, D.H. (1998). Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics, ***148***: 1269-1284.

Dietmaier, W., Wallinger, S., Bocker, T., Kullmann, F., Fishel, R., and Ruschoff, J. (1997). Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. Cancer Res., ***57***: 4749-4756.

Duckett, D.R., Bronstein, S.M., Taya, Y., and Modrich, P. (1999). hMutSalpha- and hMutLalpha-dependent phosphorylation of p53 in response to DNA methylator damage. Proc. Natl. Acad. Sci. U. S. A., ***96***: 12384-12388.

Dunlop, M.G. (1997). Colorectal cancer. B. M. J., ***314***: 1882-1885.

Dunlop, M.G., Farrington, S.M., Carothers, A.D., Wyllie, A.H., Sharp, L., Burn, J., Liu, B., Kinzler, K.W., and Vogelstein, B. (1997). Cancer risk associated with germline DNA mismatch repair gene mutations. Hum. Mol. Genet., ***6***: 105-110.

Duval, A., Gayet, J., Zhou, X.P., Iacopetta, B., Thomas, G., and Hamelin, R. (1999). Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability. Cancer Res., ***59***: 4213-4215.

Duval, A., Rolland, S., Compoint, A., Tubacher, E., Iacopetta, B., Thomas, G., and Hamelin, R. (2001). Evolution of instability at coding and non-coding repeat sequences in human MSI-H colorectal cancers. Hum. Mol. Genet., ***10***: 513-518.

Edelmann, W., Cohen, P.E., Kane, M., Lau, K., Morrow, B., Bennett, S., Umar, A., Kunkel, T., Cattoretti, G., Chaganti, R., Pollard, J.W., Kolodner, R.D., and Kucherlapati, R. (1996). Meiotic pachytene arrest in MLH1-deficient mice. Cell, ***85***: 1125-1134.

Edelmann, W., Yang, K., Umar, A., Heyer, J., Lau, K., Fan, K., Liedtke, W., Cohen, P.E., Kane, M.F., Lipford, J.R., Yu, N., Crouse, G.F., Pollard, J.W., Kunkel, T., Lipkin, M., Kolodner, R., and Kucherlapati, R. (1997). Mutation in the mismatch repair gene Msh6 causes cancer susceptibility. Cell, ***91***: 467-477.

Edelmann, W., Yang, K., Kuraguchi, M., Heyer, J., Lia, M., Kneitz, B., Fan, K., Brown, A.M., Lipkin, M., and Kucherlapati, R. (1999). Tumorigenesis in Mlh1 and Mlh1/Apc1638N mutant mice. Cancer Res., ***59*** : 1301-1307.

Edelmann, W., Umar, A., Yang, K., Heyer, J., Kucherlapati, M., Lia, M., Kneitz, B., Avdievich, E., Fan, K., Wong, E., Crouse, G., Kunkel, T., Lipkin, M., Kolodner, R.D., and Kucherlapati, R. (2000). The DNA mismatch repair genes Msh3 and Msh6 cooperate in intestinal tumor suppression. Cancer Res., ***60***: 803-807.

Eichler, E.E., Holden, J.J., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A., and Nelson, D.L. (1994). Length of uninterrupted CGG repeats determines instability in the FMR1 gene. Nat. Genet., ***8***: 88-94.

Ejima, Y., Yang, L., and Sasaki, M.S. (2000). Aberrant splicing of the ATM gene associated with shortening of the intronic mononucleotide tract in human colon tumor cell lines: a novel mutation target of microsatellite instability. Int. J. Cancer, ***86***: 262-268.

Eppert, K., Scherer, S.W., Ozcelik, H., Pirone, R., Hoodless, P., Kim, H., Tsui, L.C., Bapat, B., Gallinger, S., Andrulis, I.L., Thomsen, G.H., Wrana, J.L., and Attisano, L. (1996). MADR2 maps to 18q21 and encodes a TGFbeta-regulated MAD-related protein that is functionally mutated in colorectal carcinoma. Cell, ***86***: 543-552.

Eshleman, J.R. and Markowitz, S.D. (1996). Mismatch repair defects in human carcinogenesis. Hum. Mol. Genet., ***5***: 1489-1494.

Ewen, M.E., Sluss, H.K., Whitehouse, L.L., and Livingston, D.M. (1993). TGF beta inhibition of Cdk4 synthesis is linked to cell cycle arrest. Cell, ***74***: 1009-1020.

Farrington, S.M., Lin-Goerke, J., Ling, J., Wang, Y., Burczak, J.D., Robbins, D.J., and Dunlop, M.G. (1998). Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. Am. J. Hum. Genet., ***63***: 749-759.

Farrington, S.M. and Dunlop, M.G. (1999). Mosaicism and sporadic familial adenomatous polyposis. Am. J. Hum. Genet., ***64***: 653-658.

Farrington, S.M. and Dunlop, M.G. Familial colon cancer syndromes and their gene, In: In Genetic predisposition to cancer. Eleles, R.A. (Ed). Chapman and Hall, London Second Edition in press.

Fearnhead, N.S., Britton, M.P., and Bodmer, W.F. (2001). The ABC of APC. Hum. Mol. Genet., ***10***: 721-733.

Fearon, E.R. and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. Cell, ***61***: 759-767.

Fink, D., Aebi, S., and Howell, S.B. (1998). The role of DNA mismatch repair in drug resistance. Clin. Cancer Res., ***4***: 1-6.

Fishel, R., Lescoe, M.K., Rao, M.R., Copeland, N.G., Jenkins, N.A., Garber, J., Kane, M., and Kolodner, R. (1993). The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell, ***75***: 1027-1038.

Fishel, R. and Kolodner, R.D. (1995). Identification of mismatch repair genes and their role in the development of cancer. Curr. Opin. Genet. Dev., ***5***: 382-395.

Flores-Rozas, H. and Kolodner, R.D. (1998). The Saccharomyces cerevisiae MLH3 gene functions in MSH3-dependent suppression of frameshift mutations. Proc. Natl. Acad. Sci. U. S. A., ***95***: 12404-12409.

Franceschi, S., Favero, A., La Vecchia, C., Negri, E., Conti, E., Montella, M., Giacosa, A., Nanni, O., and Decarli, A. (1997). Food groups and risk of colorectal cancer in Italy. Int. J. Cancer, ***72***: 56-61.

Frazier, M.L., Sinicrope, F.A., Amos, C.I., Cleary, K.R., Lynch, P.M., Levin, B., and Luthra, R. (1999). Loci for efficient detection of microsatellite instability in hereditary non-polyposis colorectal cancer. Oncol. Rep., ***6***: 497-505.

Freidberg, E.C., Walker, G.C., and Siede, W. (1995). DNA repair and mutagenesis (Washigton, DC: ASM Press).

Galio, L., Bouquet, C., and Brooks, P. (1999). ATP hydrolysis-dependent formation of a dynamic ternary nucleoprotein complex with MutS and MutL. Nucleic Acids Res., ***27***: 2325-2331.

Genschel, J., Littman, S.J., Drummond, J.T., and Modrich, P. (1998). Isolation of MutSbeta from human cells and comparison of the mismatch repair specificities of MutSbeta and MutSalpha. J. Biol. Chem., ***273***: 19895-19901.

Gertig, D.M. and Hunter, D.J. (1998). Genes and environment in the etiology of colorectal cancer. Semin. Cancer Biol., ***8***: 285-298.

Giordano, M., Oefner, P.J., Underhill, P.A., Cavalli, S.L., Tosi, R., and Richiardi, P.M. (1999). Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. Genomics, ***56***: 247-253.

Giovannucci, E., Rimm, E.B., Stampfer, M.J., Colditz, G.A., Ascherio, A., and Willett, W.C. (1994). Intake of fat, meat, and fiber in relation to risk of colon cancer in men. Cancer Res., ***54***: 2390-2397.

Glickman, B.W. and Radman, M. (1980). Escherichia coli mutator mutants deficient in methylation-instructed DNA mismatch correction. Proc. Natl. Acad. Sci. U. S. A., ***77***: 1063-1067.

Goldbohm, R.A., van den Brandt, P.A., van, '., V, Brants, H.A., Dorant, E., Sturmans, F., and Hermus, R.J. (1994). A prospective cohort study on the relation between meat consumption and the risk of colon cancer. Cancer Res., ***54***: 718-723.

Goldstein, D.B. and Clark, A.G. (1995). Microsatellite variation in North American populations of Drosophila melanogaster. Nucleic Acids Res., ***23***: 3882-3886.

Goping, I.S., Gross, A., Lavoie, J.N., Nguyen, M., Jemmerson, R., Roth, K., Korsmeyer, S.J., and Shore, G.C. (1998). Regulated targeting of BAX to mitochondria. J. Cell Biol., ***143***: 207-215.

Gordenin, D.A., Kunkel, T.A., and Resnick, M.A. (1997). Repeat expansion--all in a flap? Nat. Genet., ***16***: 116-118.

Gradia, S., Acharya, S., and Fishel, R. (1997). The human mismatch recognition complex hMSH2-hMSH6 functions as a novel molecular switch. Cell, ***91***: 995-1005.

Grady, W.M., Rajput, A., Myeroff, L., Liu, D.F., Kwon, K., Willis, J., and Markowitz, S. (1998). Mutation of the type II transforming growth factor-beta receptor is coincident with the transformation of human colon adenomas to malignant carcinomas. Cancer Res., ***58***: 3101-3104.

Grady, W.M., Myeroff, L.L., Swinler, S.E., Rajput, A., Thiagalingam, S., Lutterbaugh, J.D., Neumann, A., Brattain, M.G., Chang, J., Kim, S.J., Kinzler, K.W., Vogelstein, B., Willson, J.K.,

and Markowitz, S. (1999). Mutational inactivation of transforming growth factor beta receptor type II in microsatellite stable colon cancers. Cancer Res., ***59***: 320-324.

Grilley, M., Welsh, K.M., Su, S.S., and Modrich, P. (1989). Isolation and characterisation of the Escherichia coli mutL gene product. J. Biol. Chem., ***264***: 1000-1004.

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., and Robertson, M. (1991). Identification and characterisation of the familial adenomatous polyposis coli gene. Cell, ***66***: 589-600.

Gryfe, R., Kim, H., Hsieh, E.T., Aronson, M.D., Holowaty, E.J., Bull, S.B., Redston, M., and Gallinger, S. (2000). Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. N. Engl. J. Med., ***342***: 69-77.

Guanti, G., Resta, N., Simone, C., Cariola, F., Demma, I., Fiorente, P., and Gentile, M. (2000). Involvement of PTEN mutations in the genetic pathways of colorectal cancerogenesis. Hum. Mol. Genet., ***9***: 283-287.

Gurin, C.C., Federici, M.G., Kang, L., and Boyd, J. (1999). Causes and consequences of microsatellite instability in endometrial carcinoma. Cancer Res., ***59***: 462-466.

Habraken, Y., Sung, P., Prakash, L., and Prakash, S. (1996). Binding of insertion/deletion DNA mismatches by the heterodimer of yeast mismatch repair proteins MSH2 and MSH3. Curr. Biol., ***6***: 1185-1187.

Hackman, P., Tannergard, P., Osei-mensa, S., Chen, J., Kane, M.F., Kolodener, R.D., Lambert, B., Hellgren, D., and Lindblom, A. (1997). A human compound heterozygote for two mlh1 missense mutations. Nat. Genet., ***17***: 135-136.

Haenszel, W. (1961). Cancer mortality among the foreign-born in the united states. J. Natl. Cancer Inst., ***26***: 37-132.

Haenszel, W. and Correa, P. (1971). Cancer of the colon and rectum and adenomatous polyps. A review of epidemiologic findings. Cancer, ***28***: 14-24.

Hahn, S.A., Schutte, M., Hoque, A.T., Moskaluk, C.A., da Costa, L.T., Rozenblum, E., Weinstein, C.L., Fischer, A., Yeo, C.J., Hruban, R.H., and Kern, S.E. (1996). DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. Science, ***271***: 350-353.

Hall, M.C. and Matson, S.W. (1999). The Escherichia coli MutL protein physically interacts with MutH and stimulates the MutH-associated endonuclease activity. J. Biol. Chem., ***274***: 1306-1312.

Hamilton, S.R., Liu, B., Parsons, R.E., Papadopoulos, N., Jen, J., Powell, S.M., Krush, A.J., Berk, T., Cohen, Z., and Tetu, B. (1995). The molecular basis of Turcot's syndrome. N. Engl. J. Med., ***332***: 839-847.

Han, H.J., Yanagisawa, A., Kato, Y., Park, J.G., and Nakamura, Y. (1993). Genetic instability in pancreatic cancer and poorly differentiated type of gastric cancer. Cancer Res., ***53***: 5087-5089.

Hannon, G.J. and Beach, D. (1994). p15INK4B is a potential effector of TGF-beta-induced cell cycle arrest. Nature, ***371***: 257-261.

Hardman, R.A., Afshari, C.A., and Barrett, J.C. (2001). Involvement of mammalian MLH1 in the apoptotic response to peroxide-induced oxidative stress. Cancer Res., ***61***: 1392-1397.

Harfe, B.D. and Jinks-Robertson, S. (2000). DNA mismatch repair and genetic instability. Annu. Rev. Genet., ***34***: 359-399.

Hawn, M.T., Umar, A., Carethers, J.M., Marra, G., Kunkel, T.A., Boland, C.R., and Koi, M. (1995). Evidence for a connection between the mismatch repair system and the G2 cell cycle checkpoint. Cancer Res., ***55***: 3721-3725.

Heinen, C.D., Richardson, D., White, R., and Groden, J. (1995). Microsatellite instability in colorectal adenocarcinoma cell lines that have full-length adenomatous polyposis coli protein. Cancer Res., ***55***: 4797-4799.

Heinen, C.D., Shivapurkar, N., Tang, Z., Groden, J., and Alabaster, O. (1996). Microsatellite instability in aberrant crypt foci from human colons. Cancer Res., ***56***: 5339-5341.

Hemminki, A., Peltomaki, P., Mecklin, J.P., Jarvinen, H., Salovaara, R., Nystrom-Lahti, M., de la Chapelle, A., and Aaltonen, L.A. (1994). Loss of the wild type MLH1 gene is a feature of hereditary nonpolyposis colorectal cancer. Nat. Genet., ***8***: 405-410.

Hemminki, A., Markie, D., Tomlinson, I., Avizienyte, E., Roth, S., Loukola, A., Bignell, G., Warren, W., Aminoff, M., Hoglund, P., Jarvinen, H., Kristo, P., Pelin, K., Ridanpaa, M., Salovaara, R., Toro, T., Bodmer, W., Olschwang, S., Olsen, A.S., Stratton, M.R., de la Chapelle, A., and Aaltonen, L.A. (1998). A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. Nature, ***391***: 184-187.

Henderson, S.T. and Petes, T.D. (1992). Instability of simple sequence DNA in Saccharomyces cerevisiae. Mol. Cell Biol., ***12***: 2749-2757.

Hendrich, B., Hardeland, U., Ng, H.H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature, ***401***: 301-304.

Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W., Kane, M.F., Kolodner, R.D., Vogelstein, B., Kunkel, T.A., and Baylin, S.B. (1998). Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. Proc. Natl. Acad. Sci. U. S. A., ***95***: 6870-6875.

Heyer, J., Yang, K., Lipkin, M., Edelmann, W., and Kucherlapati, R. (1999). Mouse models for colorectal cancer. Oncogene, ***18***: 5325-5333.

Hickman, M.J. and Samson, L.D. (1999). Role of DNA mismatch repair and p53 in signalling induction of apoptosis by alkylating agents. Proc. Natl. Acad. Sci. U. S. A., ***96:*** 10764-10769.

Hirst, M.C., Grewal, P.K., and Davies, K.E. (1994). Precursor arrays for triplet repeat expansion at the fragile X locus. Hum. Mol. Genet., ***3:*** 1553-1560.

Hoang, J.M., Cottu, P.H., Thuille, B., Salmon, R.J., Thomas, G., and Hamelin, R. (1997). BAT-26, an indicator of the replication error phenotype in colorectal cancers and cell lines. Cancer Res., ***57:*** 300-303.

Hoff-Olsen, P., Meling, G.I., and Olaisen, B. (1998). Variation in mutation rate and direction between tetranucleotide STR loci in human colorectal carcinomas. Ann. Hum. Genet., ***62:*** 1-7.

Homfray, T.F., Cottrell, S.E., Ilyas, M., Rowan, A., Talbot, I.C., Bodmer, W.F., and Tomlinson, I.P. (1998). Defects in mismatch repair occur after APC mutations in the pathogenesis of sporadic colorectal tumours. Hum. Mutat., ***11:*** 114-120.

Howe, J.R., Roth, S., Ringold, J.C., Summers, R.W., Jarvinen, H.J., Sistonen, P., Tomlinson, I.P., Houlston, R.S., Bevan, S., Mitros, F.A., Stone, E.M., and Aaltonen, L.A. (1998). Mutations in the SMAD4/DPC4 gene in juvenile polyposis. Science, ***280:*** 1086-1088.

Huang, J., Papadopoulos, N., McKinley, A.J., Farrington, S.M., Curtis, L.J., Wyllie, A.H., Zheng, S., Willson, J.K., Markowitz, S.D., Morin, P., Kinzler, K.W., Vogelstein, B., and Dunlop, M.G. (1996). APC mutations in colorectal tumors with mismatch repair deficiency. Proc. Natl. Acad. Sci. U. S. A., ***93:*** 9049-9054.

Inman, G.J. and Allday, M.J. (2000). Resistance to TGF-beta1 correlates with a reduction of TGF-beta type II receptor expression in Burkitt's lymphoma and Epstein-Barr virus-transformed B lymphoblastoid cell lines. J. Gen. Virol., ***81:*** 1567-1578.

International human genome sequencing consortium. (2001). Initial Sequencing and analysis of the human genome. Nature, ***409***, 860-921.

Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D., and Perucho, M. (1993). Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature, ***363:*** 558-561.

Ionov, Y., Yamamoto, H., Krajewski, S., Reed, J.C., and Perucho, M. (2000). Mutational inactivation of the proapoptotic gene BAX confers selective advantage during tumor clonal evolution. Proc. Natl. Acad. Sci. U. S. A., ***97:*** 10872-10877.

Iwao, K., Nakamori, S., Kameyama, M., Imaoka, S., Kinoshita, M., Fukui, T., Ishiguro, S., Nakamura, Y., and Miyoshi, Y. (1998). Activation of the beta-catenin gene by interstitial deletions involving exon 3 in primary colorectal carcinomas without adenomatous polyposis coli mutations. Cancer Res., ***58:*** 1021-1026.

Jass, J.R. (2000). Pathology of hereditary nonpolyposis colorectal cancer. Ann. N. Y. Acad. Sci., ***910:***62-73

Jeffreys, A.J., Neumann, R., and Wilson, V. (1990). Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell, ***60***: 473-485.

Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L., and Armour, J.A. (1994). Complex gene conversion events in germline mutation at human minisatellites. Nat. Genet., ***6***: 136-145.

Jenne, D.E., Reimann, H., Nezu, J., Friedel, W., Loff, S., Jeschke, R., Muller, O., Back, W., and Zimmer, M. (1998). Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. Nat. Genet., ***18***: 38-43.

Johnson, R.E., Kovvali, G.K., Prakash, L., and Prakash, S. (1996). Requirement of the yeast MSH3 and MSH6 genes for MSH2-dependent genomic stability. J. Biol. Chem., ***271***: 7285-7288.

Jones, P.A. and Laird, P.W. (1999). Cancer epigenetics comes of age. Nat. Genet., ***21***: 163-167.

Joslyn, G., Carlson, M., Thliveris, A., Albertsen, H., Gelbert, L., Samowitz, W., Groden, J., Stevens, J., Spirio, L., and Robertson, M. (1991). Identification of deletion mutations and three new genes at the familial polyposis locus. Cell, ***66***: 601-613.

Jurgensmeier, J.M., Xie, Z., Deveraux, Q., Ellerby, L., Bredesen, D., and Reed, J.C. (1998). Bax directly induces release of cytochrome c from isolated mitochondria. Proc. Natl. Acad. Sci. U. S. A., ***95***: 4997-5002.

Karran, P. and Marinus, M.G. (1982). Mismatch correction at O6-methylguanine residues in E. coli DNA. Nature, ***296***: 868-869.

Kashi, Y., King, D., and Soller, M. (1997). Simple sequence repeats as a source of quantitative genetic variation. Trends Genet., ***13***: 74-78.

Kim, H., Jen, J., Vogelstein, B., and Hamilton, S.R. (1994). Clinical and pathological characteristics of sporadic colorectal carcinomas with DNA replication errors in microsatellite sequences. Am. J. Pathol., ***145***: 148-156.

Kinzler, K.W., Nilbert, M.C., Su, L.K., Vogelstein, B., Bryan, T.M., Levy, D.B., Smith, K.J., Preisinger, A.C., Hedge, P., and McKechnie, D. (1991). Identification of FAP locus genes from chromosome 5q21. Science, ***253***: 661-665.

Kinzler, K.W. and Vogelstein, B. (1996). Lessons from hereditary colorectal cancer. Cell, ***87***: 159-170.

Kitaeva, M.N., Grogan, L., Williams, J.P., Dimond, E., Nakahara, K., Hausner, P., DeNobile, J.W., Soballe, P.W., and Kirsch, I.R. (1997). Mutations in beta-catenin are uncommon in colorectal cancer occurring in occasional replication error-positive tumors. Cancer Res., ***57***: 4478-4481.

Koi, M., Umar, A., Chauhan, D.P., Cherian, S.P., Carethers, J.M., Kunkel, T.A., and Boland, C.R. (1994). Human chromosome 3 corrects mismatch repair deficiency and microsatellite instability and reduces N-methyl-N'-nitro-N-nitrosoguanidine tolerance in colon tumor cells with homozygous hMLH1 mutation. Cancer Res., ***54***: 4308-4312.

Kolodner, R.D., Hall, N.R., Lipford, J., Kane, M.F., Rao, M.R., Morrison, P., Wirth, L., Finan, P.J., Burn, J., and Chapman, P. (1994). Structure of the human MSH2 locus and analysis of two Muir-Torre kindreds for msh2 mutations. Genomics, ***24***: 516-526.

Kolodner, R.D. and Marsischky, G.T. (1999). Eukaryotic DNA mismatch repair. Curr. Opin. Genet. Dev., ***9***: 89-96.

Kolodner, R.D., Tytell, J.D., Schmeits, J.L., Kane, M.F., Gupta, R.D., Weger, J., Wahlberg, S., Fox, E.A., Peel, D., Ziogas, A., Garber, J.E., Syngal, S., Anton-Culver, H., and Li, F.P. (1999). Germ-line msh6 mutations in colorectal cancer families. Cancer Res., ***59***: 5068-5074.

Konishi, M., Kikuchi-Yanoshita, R., Tanaka, K., Muraoka, M., Onda, A., Okumura, Y., Kishi, N., Iwama, T., Mori, T., Koike, M., Ushio, K., Chiba, M., Nomizu, S., Konishi, F., Utsunomiya, J., and Miyaki, M. (1996). Molecular nature of colon tumors in hereditary nonpolyposis colon cancer, familial polyposis, and sporadic colon cancer. Gastroenterology, ***111***: 307-317.

Kretzschmar, M. and Massague, J. (1998). SMADs: mediators and regulators of TGF-beta signaling. Curr. Opin. Genet. Dev., ***8***: 103-111.

Kruse, R., Lamberti, C., Wang, Y., Ruelfs, C., Bruns, A., Esche, C., Lehmann, P., Ruzicka, T., Rutten, A., Friedl, W., and Propping, P. (1996). Is the mismatch repair deficient type of Muir-Torre syndrome confined to mutations in the hMSH2 gene? Hum. Genet., ***98***: 747-750.

Kruse, R., Rutten, A., Lamberti, C., Hosseiny-Malayeri, H.R., Wang, Y., Ruelfs, C., Jungck, M., Mathiak, M., Ruzicka, T., Hartschuh, W., Bisceglia, M., Friedl, W., and Propping, P. (1998). Muir-Torre phenotype has a frequency of DNA mismatch-repair-gene mutations similar to that in hereditary nonpolyposis colorectal cancer families defined by the Amsterdam criteria. Am. J. Hum. Genet., ***63***: 63-70.

Kunst, C.B., Leeflang, E.P., Iber, J.C., Arnheim, N., and Warren, S.T. (1997). The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. J. Med. Genet., ***34***: 627-631.

La Vecchia, C., Negri, E., Franceschi, S., Conti, E., Montella, M., Giacosa, A., Falcini, A., and Decarli, A. (1997). Aspirin and colorectal cancer. Br. J. Cancer., ***76***: 675-677.

Lachance, Y., Luu-The, V., Labrie, C., Simard, J., Dumont, M., de Launoit, Y., Guerin, S., Leblanc, G., and Labrie, F. (1990). Characterization of human 3 beta-hydroxysteroid dehydrogenase/delta 5-delta 4-isomerase gene and its expression in mammalian cells. J. Biol. Chem., ***265***: 20469-20475.

Lahue, R.S., Au, K.G., and Modrich, P. (1989). DNA mismatch correction in a defined system. Science, ***245***: 160-164.

Lamers, M.H., Perrakis, A., Enzlin, J.H., Winterwerp, H.H., de Wind, N., and Sixma, T.K. (2000). The crystal structure of DNA mismatch repair protein MutS binding to a G x T mismatch. Nature, ***407***: 711-717.

Lamlum, H., Papadopoulou, A., Ilyas, M., Rowan, A., Gillet, C., Hanby, A., Talbot, I., Bodmer, W., and Tomlinson, I. (2000). APC mutations are sufficient for the growth of early colorectal adenomas. Proc. Natl. Acad. Sci. U. S. A., ***97***: 2225-2228.

Lathi, N., Peltomaki, P., Aaltonen, L.A., and de la Chapelle, A. (1994). Genetic and genealogic study of 33 finnish HNPCC kindreds. Am. J. Hum. Genet., ***55***: A535

Laurent-Puig, P., Beroud, C., and Soussi, T. (1998). APC gene: database of germline and somatic mutations in human tumors and cell lines. Nucleic Acids Res., ***26***: 269-270.

Le Marchand, L., Wilkens, L.R., Kolonel, L.N., Hankin, J.H., and Lyu, L.C. (1997). Associations of sedentary lifestyle, obesity, smoking, alcohol use, and diabetes with the risk of colorectal cancer. Cancer Res., ***57***: 4787-4794.

Leach, F.S., Nicolaides, N.C., Papadopoulos, N., Liu, B., Jen, J., Parsons, R., Peltomaki, P., Sistonen, P., Aaltonen, L.A., and Nystrom-Lahti, M. (1993). Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell., ***75***: 1215-1225.

Leibovitz, A., Stinson, J.C., McCombs, W.B., McCoy, C.E., Mazur, K.C., and Mabry, N.D. (1976). Classification of human colorectal adenocarcinoma cell lines. Cancer Res., ***36***: 4562-4569.

Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1998). Genetic instabilities in human cancers. Nature, ***396***: 643-649.

Levi, F., Pasche, C., La Vecchia, C., Lucchini, F., and Franceschi, S. (1999). Food groups and colorectal cancer risk. Br. J. Cancer., ***79***: 1283-1287.

Levinson, G. and Gutman, G.A. (1987). High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in Escherichia coli K-12. Nucleic Acids Res., ***15***: 5323-5338.

Li, G.M. and Modrich, P. (1995). Restoration of mismatch repair to nuclear extracts of H6 colorectal tumor cells by a heterodimer of human MutL homologs. Proc. Natl. Acad. Sci. U. S. A., ***92***: 1950-1954.

Liaw, D., Marsh, D.J., Li, J., Dahia, P.L., Wang, S.I., Zheng, Z., Bose, S., Call, K.M., Tsou, H.C., Peacocke, M., Eng, C., and Parsons, R. (1997). Germline mutations of the PTEN gene in Cowden disease, an inherited breast and thyroid cancer syndrome. Nat. Genet., ***16***: 64-67.

Lin, H.Y., Wang, X.F., Ng-Eaton, E., Weinberg, R.A., and Lodish, H.F. (1992). Expression cloning of the TGF-beta type II receptor, a functional transmembrane serine/threonine kinase. Cell, ***68***: 775-785.

Lindblom, A., Tannergard, P., Werelius, B., and Nordenskjold, M. (1993). Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. Nat. Genet., ***5***: 279-282.

Lipkin, S.M., Wang, V., Jacoby, R., Banerjee-Basu, S., Baxevanis, A.D., Lynch, H.T., Elliott, R.M., and Collins, F.S. (2000) MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nat. Genet., ***24***: 27-35.

Liu, B., Farrington, S.M., Petersen, G.M., Hamilton, S.R., Parsons, R., Papadopoulos, N., Fujiwara, T., Jen, J., Kinzler, K.W., Wyllie, A.H., Vogelstein, B., and Dunlop, M. (1995a). Genetic instability occurs in the majority of young patients with colorectal cancer. Nat. Med., ***1***: 348-352.

Liu, B., Nicolaides, N.C., Markowitz, S., Willson, J.K., Parsons, R.E., Jen, J., Papadopolous, N., Peltomaki, P., de la Chapelle, A., and Hamilton, S.R. (1995b). Mismatch repair gene defects in sporadic colorectal cancers with microsatellite instability. Nat. Genet., ***9***: 48-55.

Liu, B., Parsons, R., Papadopoulos, N., Nicolaides, N.C., Lynch, H.T., Watson, P., Jass, J.R., Dunlop, M., Wyllie, A., Peltomaki, P., de la Chapelle, A., Hamilton, S.R., Vogelstein, B., and Kinzler, K.W. (1996). Analysis of mismatch repair genes in hereditary non-polyposis colorectal cancer patients. Nat. Med., ***2***: 169-174.

Liu, W., Dong, X., Mai, M., Seelan, R.S., Taniguchi, K., Krishnadath, K.K., Halling, K.C., Cunningham, J.M., Qian, C., Christensen, E., Roche, P.C., Smith, D.I., and Thibodeau, S.N. (2000). Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling. Nat. Genet., ***26***: 146-147.

Loeb, L.A., Springgate, C.F., and Battula, N. (1974). Errors in DNA replication as a basis of malignant changes. Cancer Res., ***34***: 2311-2321.

Loeb, L.A. (1994). Microsatellite instability: marker of a mutator phenotype in cancer. Cancer Res., ***54***: 5059-5063.

Loeb, L.A. (2001). A mutator phenotype in cancer. Cancer Res., ***61***: 3230-3239.

Longnecker, M.P., Orza, M.J., Adams, M.E., Vioque, J., and Chalmers, T.C. (1990). A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer. Cancer Causes Control, ***1***: 59-68.

Lothe, R.A., Peltomaki, P., Meling, G.I., Aaltonen, L.A., Nystrom-Lahti, M., Pylkkanen, L., Heimdal, K., Andersen, T.I., Moller, P., and Rognum, T.O. (1993). Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. Cancer Res., ***53***: 5849-5852.

Loukola, A., Vilkki, S., Singh, J., Launonen, V., and Aaltonen, L.A. (2000). Germline and somatic mutation analysis of MLH3 in MSI-positive colorectal cancer. Am. J. Pathol., ***157***: 347-352.

Lu, S.L., Zhang, W.C., Akiyama, Y., Nomizu, T., and Yuasa, Y. (1996). Genomic structure of the transforming growth factor beta type II receptor gene and its mutations in hereditary nonpolyposis colorectal cancers. Cancer Res., ***56***: 4595-4598.

Lu, S.L., Kawabata, M., Imamura, T., Akiyama, Y., Nomizu, T., Miyazono, K., and Yuasa, Y. (1998). HNPCC associated with germline mutation in the TGF-beta type II receptor gene. Nat Genet., ***19***: 17-18.

Lynch, H.T., Shaw, M.W., Magnuson, C.W., Larsen, A.L., and Krush, A.J. (1966). Hereditary factors in cancer. Study of two large midwestern kindreds. Arch. Intern. Med., ***117***: 206-212.

Lynch, H.T. and Krush, A.J. (1971). Cancer family "G" revisited: 1895-1970. Cancer, ***27***: 1505-1511.

Lynch, H.T., Kimberling, W., Albano, W.A., Lynch, J.F., Biscone, K., Schuelke, G.S., Sandberg, A.A., Lipkin, M., Deschner, E.E., and Mikol, Y.B. (1985). Hereditary nonpolyposis colorectal cancer (Lynch syndromes I and II). I. Clinical description of resource. Cancer, ***56***: 934-938.

Lynch, H.T. and Smyrk, T. (1996). Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. Cancer, ***78***: 1149-1167.

Lynch, E.D., Ostermeyer, E.A., Lee, M.K., Arena, J.F., Ji, H., Dann, J., Swisshelm, K., Suchard, D., MacLeod, P.M., Kvinnsland, S., Gjertsen, B.T., Heimdal, K., Lubs, H., Moller, P., and King, M.C. (1997). Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. Am. J. Hum. Genet., ***61***: 1254-1260.

Lynch, H.T. and de la Chapelle, A. (1999). Genetic susceptibility to non-polyposis colorectal cancer. J. Med. Genet., ***36***: 801-818.

Ma, A.H., Xia, L., Littman, S.J., Swinler, S., Lader, G., Polinkovsky, A., Olechnowicz, J., Kasturi, L., Lutterbaugh, J., Modrich, P., Veigl, M.L., Markowitz, S.D., and Sedwick, W.D. (2000). Somatic mutation of hPMS2 as a possible cause of sporadic human colon cancer with microsatellite instability. Oncogene, ***19***: 2249-2256.

Macleod, K. (2000). Tumor suppressor genes. Curr. Opin. Genet. Dev., ***10***: 81-93.

Mai, M., Qian, C., Yokomizo, A., Smith, D.I., and Liu, W. (1999). Cloning of the human homolog of conductin (AXIN2), a gene mapping to chromosome 17q23-q24. Genomics, ***55***: 341-344.

Malkhosyan, S., Rampino, N., Yamamoto, H., and Perucho, M. (1996). Frameshift mutator mutations. Nature, ***382***: 499-500.

Mann, B., Gelos, M., Siedow, A., Hanski, M.L., Gratchev, A., Ilyas, M., Bodmer, W.F., Moyer, M.P., Riecken, E.O., Buhr, H.J., and Hanski, C. (1999). Target genes of beta-catenin-T cell-factor/lymphoid-enhancer-factor signaling in human colorectal carcinomas. Proc. Natl. Acad. Sci. U. S. A., ***96***: 1603-1608.

Markowitz, S., Wang, J., Myeroff, L., Parsons, R., Sun, L., Lutterbaugh, J., Fan, R.S., Zborowska, E., Kinzler, K.W., and Vogelstein, B. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science, ***268***: 1336-1338.

Markowitz, S. (2000). TGF-beta receptors and DNA repair genes, coupled targets in a pathway of human colon carcinogenesis. Biochim. Biophys. Acta., ***1470***: M13-M20.

Marsischky, G.T., Filosi, N., Kane, M.F., and Kolodner, R. (1996). Redundancy of Saccharomyces cerevisiae MSH3 and MSH6 in MSH2-dependent mismatch repair. Genes Dev., ***10***: 407-420.

Marzo, I., Brenner, C., Zamzami, N., Jurgensmeier, J.M., Susin, S.A., Vieira, H.L., Prevost, M.C., Xie, Z., Matsuyama, S., Reed, J.C., and Kroemer, G. (1998). Bax and adenine nucleotide translocator cooperate in the mitochondrial control of apoptosis. Science, ***281***: 2027-2031.

Massague, J. (1996). TGFbeta signaling: receptors, transducers, and Mad proteins. Cell, ***85***: 947-950.

Matsui, T., Hayashi, N., Yao, K., Yao, T., Takenaka, K., Hoashi, T., Takemura, S., Iwashita, A., Tanaka, A., and Koga, M. (1998). A father and son with Turcot's syndrome: evidence for autosomal dominant inheritance: report of two cases. Dis. Colon Rectum, ***41***: 797-801.

May, C.A., Jeffreys, A.J., and Armour, J.A. (1996). Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). Hum. Mol. Genet., ***5***: 1823-1833.

McMichael, A.J. and Giles, G.G. (1988). Cancer in migrants to Australia: extending the descriptive epidemiological data. Cancer Res., ***48***: 751-756.

Mecklin, J.P., Jarvinen, H.J., Hakkiluoto, A., Hallikas, H., Hiltunen, K.M., Harkonen, N., Kellokumpu, I., Laitinen, S., Ovaska, J., and Tulikoura, J. (1995). Frequency of hereditary nonpolyposis colorectal cancer. A prospective multicenter study in Finland. Dis. Colon Rectum, ***38***: 588-593.

Merlo, A., Mabry, M., Gabrielson, E., Vollmer, R., Baylin, S.B., and Sidransky, D. (1994). Frequent microsatellite instability in primary small cell lung cancer. Cancer Res., ***54***: 2098-2101.

Mirabelli-Primdahl, L., Gryfe, R., Kim, H., Millar, A., Luceri, C., Dale, D., Holowaty, E., Bapat, B., Gallinger, S., and Redston, M. (1999). Beta-catenin mutations are specific for colorectal carcinomas with microsatellite instability but occur in endometrial carcinomas irrespective of mutator pathway. Cancer Res., ***59***: 3346-3351.

Mironov, N., Jansen, L.A., Zhu, W.B., Aguelon, A.M., Reguer, G., and Yamasaki, H. (1999). A novel sensitive method to detect frameshift mutations in exonic repeat sequences of cancer-related genes. Carcinogenesis, ***20***: 2189-2192.

Miyaki, M., Konishi, M., Kikuchi-Yanoshita, R., Enomoto, M., Igari, T., Tanaka, K., Muraoka, M., Takahashi, H., Amada, Y., and Fukayama, M. (1994). Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. Cancer Res., ***54***: 3011-3020.

Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M., and Mori, T. (1997a). Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. Nat. Genet., ***17***: 271-272.

Miyaki, M., Nishio, J., Konishi, M., Kikuchi-Yanoshita, R., Tanaka, K., Muraoka, M., Nagato, M., Chong, J.M., Koike, M., Terada, T., Kawahara, Y., Fukutome, A., Tomiyama, J., Chuganji, Y., Momoi, M., and Utsunomiya, J. (1997b). Drastic genetic instability of tumors and normal tissues in Turcot syndrome. Oncogene, ***15***: 2877-2881.

Miyaki, M., Iijima, T., Kimura, J., Yasuno, M., Mori, T., Hayashi, Y., Koike, M., Shitara, N., Iwama, T., and Kuroki, T. (1999). Frequent mutation of beta-catenin and APC genes in primary colorectal tumors from patients with hereditary nonpolyposis colorectal cancer. Cancer Res., ***59*** : 4506-4509.

Miyashita, T. and Reed, J.C. (1995). Tumor suppressor p53 is a direct transcriptional activator of the human bax gene. Cell, ***80***: 293-299.

Miyoshi, Y., Nagase, H., Ando, H., Horii, A., Ichii, S., Nakatsuru, S., Aoki, T., Miki, Y., Mori, T., and Nakamura, Y. (1992). Somatic mutations of the APC gene in colorectal tumors: mutation cluster region in the APC gene. Hum. Mol. Genet., ***1***: 229-233.

Modrich, P. (1991). Mechanisms and biological effects of mismatch repair. Annu. Rev. Genet., ***25:229-53***: 229-253.

Modrich, P. (1994). Mismatch repair, genetic stability, and cancer. Science, ***266***: 1959-1960.

Modrich, P. and Lahue, R. (1996). Mismatch repair in replication fidelity, genetic recombination, and cancer biology. Annu. Rev. Biochem., ***65:101-33***: 101-133.

Monckton, D.G., Wong, L.J., Ashizawa, T., and Caskey, C.T. (1995). Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. Hum. Mol. Genet., ***4***: 1-8.

Monckton, D.G., Cayuela, M.L., Gould, F.K., Brock, G.J., Silva, R., and Ashizawa, T. (1999). Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. Hum. Mol. Genet., ***8***: 2473-2478.

Morin, P.J., Sparks, A.B., Korinek, V., Barker, N., Clevers, H., Vogelstein, B., and Kinzler, K.W. (1997). Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta-catenin or APC. Science, ***275***: 1787-1790.

Mornet, E., Chateau, C., Hirst, M.C., Thepot, F., Taillandier, A., Cibois, O., and Serre, J.L. (1996). Analysis of germline variation at the FMR1 CGG repeat shows variation in the normal-premutated borderline range. Hum. Mol. Genet., ***5***: 821-825.

Moses, H.L., Yang, E.Y., and Pietenpol, J.A. (1990). TGF-beta stimulation and inhibition of cell proliferation: new mechanistic insights. Cell, ***63***: 245-247.

Moslein, G., Tester, D.J., Lindor, N.M., Honchel, R., Cunningham, J.M., French, A.J., Halling, K.C., Schwab, M., Goretzki, P., and Thibodeau, S.N. (1996). Microsatellite instability and mutation analysis of hMSH2 and hMLH1 in patients with sporadic, familial and hereditary colorectal cancer. Hum. Mol. Genet., ***5***: 1245-1252.

Muir, E.G., Bell, A.J., and Barlow, K.A. (1967). Multiple primary carcinomata of the colon, duodenum, and larynx associated with kerato-acanthomata of the face. Br. J. Surg., ***54***: 191-195.

Munemitsu, S., Albert, I., Souza, B., Rubinfeld, B., and Polakis, P. (1995). Regulation of intracellular beta-catenin levels by the adenomatous polyposis coli (APC) tumor-suppressor protein. Proc. Natl. Acad. Sci. U. S. A., ***92***: 3046-3050.

Muscat, J.E., Stellman, S.D., and Wynder, E.L. (1994). Nonsteroidal antiinflammatory drugs and colorectal cancer. Cancer, ***74***: 1847-1854.

Nakao, A., Imamura, T., Souchelnytskyi, S., Kawabata, M., Ishisaki, A., Oeda, E., Tamaki, K., Hanai, J., Heldin, C.H., Miyazono, K., and ten Dijke, P. (1997). TGF-beta receptor-mediated signalling through Smad2, Smad3 and Smad4. EMBO J., ***16***: 5353-5362.

Narayanan, L., Fritzell, J.A., Baker, S.M., Liskay, R.M., and Glazer, P.M. (1997). Elevated levels of mutation in multiple tissues of mice deficient in the DNA mismatch repair gene Pms2. Proc. Natl. Acad. Sci. U. S. A., ***94***: 3122-3127.

Nicolaides, N.C., Papadopoulos, N., Liu, B., Wei, Y.F., Carter, K.C., Ruben, S.M., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., and Fraser, C.M. (1994). Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature, ***371***: 75-80.

Nicolaides, N.C., Littman, S.J., Modrich, P., Kinzler, K.W., and Vogelstein, B. (1998). A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. Mol. Cell Biol., ***18***: 1635-1641.

Nishisho, I., Nakamura, Y., Miyoshi, Y., Miki, Y., Ando, H., Horii, A., Koyama, K., Utsunomiya, J., Baba, S., and Hedge, P. (1991). Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients. Science, ***253***: 665-669.

Nowell, P.C. (1976). The clonal evolution of tumor cell populations. Science, ***194***: 23-28.

Obmolova, G., Ban, C., Hsieh, P., and Yang, W. (2000). Crystal structures of mismatch repair protein MutS and its complex with a substrate DNA. Nature, ***407***: 703-710.

Oki, E., Oda, S., Maehara, Y., and Sugimachi, K. (1999). Mutated gene-specific phenotypes of dinucleotide repeat instability in human colorectal carcinoma cell lines deficient in DNA mismatch repair. Oncogene, ***18***: 2143-2147.

Olschwang, S., Hamelin, R., Laurent-Puig, P., Thuille, B., De Rycke, Y., Li, Y.J., Muzeau, F., Girodet, J., Salmon, R.J., and Thomas, G. (1997). Alternative genetic pathways in colorectal carcinogenesis. Proc. Natl. Acad. Sci. U. S. A., ***94***: 12122-12127.

Olschwang, S., Serova-Sinilnikova, O.M., Lenoir, G.M., and Thomas, G. (1998). PTEN germ-line mutations in juvenile polyposis coli. Nat. Genet., ***18***: 12-14.

Oltvai, Z.N., Milliman, C.L., and Korsmeyer, S.J. (1993). Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. Cell, ***74***: 609-619.

Orford, K., Crockett, C., Jensen, J.P., Weissman, A.M., and Byers, S.W. (1997). Serine phosphorylation-regulated ubiquitination and degradation of beta-catenin. J. Biol. Chem., ***272***: 24735-24738.

Orth, K., Hung, J., Gazdar, A., Bowcock, A., Mathis, J.M., and Sambrook, J. (1994). Genetic instability in human ovarian cancer cell lines. Proc. Natl. Acad. Sci. U. S. A., ***91***: 9495-9499.

Ottini, L., Palli, D., Falchetti, M., D'Amico, C., Noviello, C., Saieva, C., Russo, A., Corsi, A.M., Masala, G., Paglierani, M., Vezzosi, V., Bianchi, S., and Mariani-Costantini, R. (2000). Instability at dinucleotide and trinucleotide repeats in breast cancer. Int. J. Oncol., ***17***: 819-826.

Ouyang, H., Shiwaku, H.O., Hagiwara, H., Miura, K., Abe, T., Kato, Y., Ohtani, H., Shiiba, K., Souza, R.F., Meltzer, S.J., and Horii, A. (1997). The insulin-like growth factor II receptor gene is mutated in genetically unstable cancers of the endometrium, stomach, and colorectum. Cancer Res., ***57***: 1851-1854.

Ouyang, H., Furukawa, T., Abe, T., Kato, Y., and Horii, A. (1998). The BAX gene, the promoter of apoptosis, is mutated in genetically unstable cancers of the colorectum, stomach, and endometrium. Clin. Cancer Res., ***4***: 1071-1074.

Palombo, F., Gallinari, P., Iaccarino, I., Lettieri, T., Hughes, M., D'Arrigo, A., Truong, O., Hsuan, J.J., and Jiricny, J. (1995). GTBP, a 160-kilodalton protein essential for mismatch-binding activity in human cells. Science, ***268***: 1912-1914.

Palombo, F., Iaccarino, I., Nakajima, E., Ikejima, M., Shimada, T., and Jiricny, J. (1996). hMutSbeta, a heterodimer of hMSH2 and hMSH3, binds to insertion/deletion loops in DNA. Curr. Biol., ***6***: 1181-1184.

Papadopoulos, N., Nicolaides, N.C., Wei, Y.F., Ruben, S.M., Carter, K.C., Rosen, C.A., Haseltine, W.A., Fleischmann, R.D., Fraser, C.M., and Adams, M.D. (1994). Mutation of a mutL homolog in hereditary colon cancer. Science, ***263***: 1625-1629.

Papadopoulos, N., Nicolaides, N.C., Liu, B., Parsons, R., Lengauer, C., Palombo, F., D'Arrigo, A., Markowitz, S., Willson, J.K., and Kinzler, K.W. (1995). Mutations of GTBP in genetically unstable cells. Science, ***268***: 1915-1917.

Parkin, D.M., Pisani, P., and Ferlay, J. (1999). Estimates of the worldwide incidence of 25 major cancers in 1990. Int. J. Cancer, ***80***: 827-841.

Parsons, R., Li, G.M., Longley, M.J., Fang, W.H., Papadopoulos, N., Jen, J., de la Chapelle, A., Kinzler, K.W., Vogelstein, B., and Modrich, P. (1993). Hypermutability and mismatch repair deficiency in RER+ tumor cells. Cell, ***75:*** 1227-1236.

Parsons, R., Li, G.M., Longley, M., Modrich, P., Liu, B., Berk, T., Hamilton, S.R., Kinzler, K.W., and Vogelstein, B. (1995a). Mismatch repair deficiency in phenotypically normal human cells. Science, ***268:*** 738-740.

Parsons, R., Myeroff, L.L., Liu, B., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (1995b). Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. Cancer Res., ***55:*** 5548-5550.

Pawlowski, J. and Kraft, A.S. (2000). Bax-induced apoptotic cell death. Proc. Natl. Acad. Sci. U. S. A., ***97:*** 529-531.

Pedroni, M., Sala, E., Scarselli, A., Borghi, F., Menigatti, M., Benatti, P., Percesepe, A., Rossi, G., Foroni, M., Losi, L., Di Gregorio, C., De Pol, A., Nascimbeni, R., Di Betta, E., Salerni, B., de Leon, M.P., and Roncucci, L. (2001). Microsatellite instability and mismatch-repair protein expression in hereditary and sporadic colorectal carcinogenesis. Cancer Res., ***61:*** 896-899.

Peinado, M.A., Malkhosyan, S., Velazquez, A., and Perucho, M. (1992). Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. Proc. Natl. Acad. Sci. U. S. A., ***89:*** 10065-10069.

Peltomaki, P., Aaltonen, L.A., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Green, J.S., Jass, J.R., Weber, J.L., and Leach, F.S. (1993). Genetic mapping of a locus predisposing to human colorectal cancer. Science, ***260:*** 810-812.

Peltomaki, P. and Vasen, H.F. (1997). Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. Gastroenterology, ***113:*** 1146-1158.

Peltomaki, P. (2001). Deficient DNA mismatch repair: a common etiologic factor for colon cancer. Hum. Mol. Genet., ***10:*** 735-740.

Perucho, M. (1996). Microsatellite instability: the mutator that mutates the other mutator. Nat. Med., ***2:*** 630-631.

Perucho, M. (1999). Correspondence re: C.R. Boland et al., A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res., 58: 5248-5257, 1998 [letter]. Cancer Res., ***59:*** 249-256.

Petes, T.D., Greenwell, P.W., and Dominska, M. (1997). Stabilization of microsatellite sequences by variant repeats in the yeast Saccharomyces cerevisiae. Genetics, ***146:*** 491-498.

Pisani, P., Parkin, D.M., Bray, F., and Ferlay, J. (1999). Estimates of the worldwide mortality from 25 cancers in 1990. Int. J. Cancer, ***83:*** 18-29.

Potter, J.D. (1999). Colorectal cancer: molecules and populations. J. Natl. Cancer Inst., ***91***: 916-932.

Powell, S.M., Zilz, N., Beazer-Barclay, Y., Bryan, T.M., Hamilton, S.R., Thibodeau, S.N., Vogelstein, B., and Kinzler, K.W. (1992). APC mutations occur early during colorectal tumorigenesis. Nature, ***359*** : 235-237.

Prolla, T.A., Pang, Q., Alani, E., Kolodner, R.D., and Liskay, R.M. (1994). MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science, ***265***: 1091-1093.

Prolla, T.A. (1998a). DNA mismatch repair and cancer. Curr. Opin. Cell Biol., ***10***: 311-316.

Prolla, T.A., Baker, S.M., Harris, A.C., Tsao, J.L., Yao, X., Bronner, C.E., Zheng, B., Gordon, M., Reneker, J., Arnheim, N., Shibata, D., Bradley, A., and Liskay, R.M. (1998b). Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. Nat. Genet., ***18***: 276-279.

Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C., and Perucho, M. (1997). Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. Science, ***275***: 967-969.

Raschle, M., Marra, G., Nystrom-Lahti, M., Schar, P., and Jiricny, J. (1999). Identification of hMutLbeta, a heterodimer of hMLH1 and hPMS1. J. Biol. Chem., ***274***: 32368-32375.

Reitmair, A.H., Schmits, R., Ewel, A., Bapat, B., Redston, M., Mitri, A., Waterhouse, P., Mittrucker, H.W., Wakeham, A., and Liu, B. (1995). MSH2 deficient mice are viable and susceptible to lymphoid tumours. Nat. Genet., ***11***: 64-70.

Reitmair, A.H., Redston, M., Cai, J.C., Chuang, T.C., Bjerknes, M., Cheng, H., Hay, K., Gallinger, S., Bapat, B., and Mak, T.W. (1996). Spontaneous intestinal carcinomas and skin neoplasms in Msh2-deficient mice. Cancer Res., ***56***: 3842-3849.

Reitmair, A.H., Risley, R., Bristow, R.G., Wilson, T., Ganesh, A., Jang, A., Peacock, J., Benchimol, S., Hill, R.P., Mak, T.W., Fishel, R., and Meuth, M. (1997). Mutator phenotype in Msh2-deficient murine embryonic fibroblasts. Cancer Res., ***57***: 3765-3771.

Ricciardone, M.D., Ozcelik, T., Cevher, B., Ozdag, H., Tuncer, M., Gurgey, A., Uzunalimoglu, O., Cetinkaya, H., Tanyeli, A., Erken, E., and Ozturk, M. (1999). Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. Cancer Res., ***59***: 290-293.

Riccio, A., Aaltonen, L.A., Godwin, A.K., Loukola, A., Percesepe, A., Salovaara, R., Masciullo, V., Genuardi, M., Paravatou-Petsotas, M., Bassi, D.E., Ruggeri, B.A., Klein-Szanto, A.J., Testa, J.R., Neri, G., and Bellacosa, A. (1999). The DNA repair gene MBD4 (MED1) is mutated in human carcinomas with microsatellite instability. Nat. Genet., ***23***: 266-268.

Risinger, J.I., Berchuck, A., Kohler, M.F., Watson, P., Lynch, H.T., and Boyd, J. (1993). Genetic instability of microsatellites in endometrial carcinoma. Cancer Res., ***53***: 5100-5103.

Risinger, J.I., Barrett, J.C., Watson, P., Lynch, H.T., and Boyd, J. (1996). Molecular genetic evidence of the occurrence of breast cancer as an integral tumor in patients with the hereditary nonpolyposis colorectal carcinoma syndrome. Cancer, ***77***: 1836-1843.

Rodriguez-Bigas, M.A., Boland, C.R., Hamilton, S.R., Henson, D.E., Jass, J.R., Khan, P.M., Lynch, H., Perucho, M., Smyrk, T., Sobin, L., and Srivastava, S. (1997). A National Cancer Institute Workshop on Hereditary Nonpolyposis Colorectal Cancer Syndrome: meeting highlights and Bethesda guidelines. J. Natl. Cancer Inst., ***89***: 1758-1762.

Rubinfeld, B., Souza, B., Albert, I., Muller, O., Chamberlain, S.H., Masiarz, F.R., Munemitsu, S., and Polakis, P. (1993). Association of the APC gene product with beta-catenin. Science, ***262***: 1731-1734.

Rubinfeld, B., Albert, I., Porfiri, E., Fiol, C., Munemitsu, S., and Polakis, P. (1996). Binding of GSK3beta to the APC-beta-catenin complex and regulation of complex assembly. Science, ***272***: 1023-1026.

Rubinfeld, B., Albert, I., Porfiri, E., Munemitsu, S., and Polakis, P. (1997). Loss of beta-catenin regulation by the APC tumor suppressor protein correlates with loss of structure due to common somatic mutations of the gene. Cancer Res., ***57***: 4624-4630.

Salahshor, S., Kressner, U., Pahlman, L., Glimelius, B., Lindmark, G., and Lindblom, A. (1999). Colorectal cancer with and without microsatellite instability involves different genes. Genes Chromosomes.Cancer, ***26***: 247-252.

Samowitz, W.S., Slattery, M.L., Potter, J.D., and Leppert, M.F. (1999). BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. Am. J. Pathol., ***154***: 1637-1641.

Sankila, R., Aaltonen, L.A., Jarvinen, H.J., and Mecklin, J.P. (1996). Better survival rates in patients with MLH1-associated hereditary colorectal cancer. Gastroenterology, ***110***: 682-687.

Shibata, D., Peinado, M.A., Ionov, Y., Malkhosyan, S., and Perucho, M. (1994). Genomic instability in repeated sequences is an early somatic event in colorectal tumorigenesis that persists after transformation. Nat. Genet., ***6***: 273-281.

Sia, E.A., Kokoska, R.J., Dominska, M., Greenwell, P., and Petes, T.D. (1997). Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol. Cell Biol., ***17***: 2851-2858.

Sieber, O.M., Tomlinson, I.P., and Lamlum, H. (2000). The adenomatous polyposis coli (APC) tumour suppressor--genetics, function and disease. Mol. Med. Today, ***6***: 462-469.

Snow, K., Tester, D.J., Kruckeberg, K.E., Schaid, D.J., and Thibodeau, S.N. (1994). Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. Hum. Mol. Genet., ***3***: 1543-1551.

Souza, R.F., Appel, R., Yin, J., Wang, S., Smolinski, K.N., Abraham, J.M., Zou, T.T., Shi, Y.Q., Lei, J., Cottrell, J., Cymes, K., Biden, K., Simms, L., Leggett, B., Lynch, P.M., Frazier, M., Powell, S.M., Harpaz, N., Sugimura, H., Young, J., and Meltzer, S.J. (1996). Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours. Nat. Genet., ***14***: 255-257.

Sparks, A.B., Morin, P.J., Vogelstein, B., and Kinzler, K.W. (1998). Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. Cancer Res., ***58***: 1130-1134.

Spitzner, J.R., Chung, I.K., and Muller, M.T. (1990). Eukaryotic topoisomerase II preferentially cleaves alternating purine-pyrimidine repeats. Nucleic Acids Res., ***18***: 1-11.

Stark, L.A., Din, F.V., Zwacka, R.M., and Dunlop, M.G. (2001). Aspirin-induced activation of the NF-kappaB signaling pathway: a novel mechanism for aspirin-mediated apoptosis in colon cancer cells. FASEB J., ***15***: 1273-1275.

Stemmermann, G.N. (1970). Patterns of disease among Japanese living in Hawaii. Arch. Environ. Health., ***20***: 266-273.

Strand, M., Prolla, T.A., Liskay, R.M., and Petes, T.D. (1993). Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature, ***365***: 274-276.

Sturzeneker, R., Bevilacqua, R.A., Haddad, L.A., Simpson, A.J., and Pena, S.D. (2000). Microsatellite instability in tumors as a model to study the process of microsatellite mutations. Hum. Mol. Genet., ***9***: 347-352.

Su, S.S. and Modrich, P. (1986). Escherichia coli mutS-encoded protein binds to mismatched DNA base pairs. Proc. Natl. Acad. Sci. U. S. A., ***83***: 5057-5061.

Su, L.K., Vogelstein, B., and Kinzler, K.W. (1993). Association of the APC tumor suppressor protein with catenins. Science, ***262***: 1734-1737.

Takenoshita, S., Tani, M., Nagashima, M., Hagiwara, K., Bennett, W.P., Yokota, J., and Harris, C.C. (1997). Mutation analysis of coding sequences of the entire transforming growth factor beta type II receptor gene in sporadic human colon cancer using genomic DNA and intron primers. Oncogene, ***14***: 1255-1258.

Taliani, M.R., Roberts, S.C., Dukek, B.A., Pruthi, R.K., Nichols, W.L., and Heit, J.A. (2001). Sensitivity and specificity of denaturing high-pressure liquid chromatography for unknown protein C gene mutations. Genet. Test., ***5***: 39-44.

Tetsu, O. and McCormick, F. (1999). Beta-catenin regulates expression of cyclin D1 in colon carcinoma cells. Nature, ***398***: 422-426.

Thibodeau, S.N., Bren, G., and Schaid, D. (1993). Microsatellite instability in cancer of the proximal colon. Science, ***260***: 816-819.

Thibodeau, S.N., French, A.J., Cunningham, J.M., Tester, D., Burgart, L.J., Roche, P.C., McDonnell, S.K., Schaid, D.J., Vockley, C.W., Michels, V.V., Farr, G.H.J., and O'Connell, M.J. (1998). Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Res., ***58***: 1713-1718.

Thornborrow, E.C. and Manfredi, J.J. (1999). One mechanism for cell type-specific regulation of the bax promoter by the tumor suppressor p53 is dictated by the p53 response element. J. Biol. Chem., ***274***: 33747-33756.

Tomlinson, I.P., Novelli, M.R., and Bodmer, W.F. (1996). The mutation rate and cancer. Proc. Natl. Acad. Sci. U. S. A., ***93***: 14800-14803.

Tomlinson, I. and Bodmer, W. (1999). Selection, the Mutation rate and cancer: ensuring that the tail does not wag the dog. Nat. Med., ***5***: 11-12.

Torre, D. (1968). Multiple sebaceous tumors. Arch. Dermatol., ***98***: 549-551.

Toth, G., Gaspari, Z., and Jurka, J. (2000). Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res., ***10***: 967-981.

Trock, B., Lanza, E., and Greenwald, P. (1990). Dietary fiber, vegetables, and colon cancer: critical review and meta-analyses of the epidemiologic evidence. J. Natl. Cancer Inst., ***82***: 650-661.

Turcot, J.T., Despres, J.-P., and St.Pierre, F. (1959). Malignant tumours of the central nervous system associated with familial polyposis of the colon: report of two cases. Dis. Colon Rectum, ***2***: 465-468.

Umar, A., Risinger, J.I., Glaab, W.E., Tindall, K.R., Barrett, J.C., and Kunkel, T.A. (1998). Functional overlap in mismatch repair by human MSH3 and MSH6. Genetics, ***148***: 1637-1646.

Vasen, H.F., Mecklin, J.P., Khan, P.M., and Lynch, H.T. (1991). The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Dis. Colon Rectum, ***34***: 424-425.

Vasen, H.F., Watson, P., Mecklin, J.P., and Lynch, H.T. (1999). New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology, ***116***: 1453-1456.

Veigl, M.L., Kasturi, L., Olechnowicz, J., Ma, A.H., Lutterbaugh, J.D., Periyasamy, S., Li, G.M., Drummond, J., Modrich, P.L., Sedwick, W.D., and Markowitz, S.D. (1998). Biallelic inactivation of hMLH1 by epigenetic gene silencing, a novel mechanism causing human MSI cancers. Proc. Natl. Acad. Sci. U. S. A., ***95***: 8698-8702.

Vilkki, S., Tsao, J.L., Loukola, A., Poyhonen, M., Vierimaa, O., Herva, R., Aaltonen, L.A., and Shibata, D. (2001). Extensive somatic microsatellite mutations in normal human tissue. Cancer Res., ***61***: 4541-4544.

Vogelstein, B., Fearon, E.R., Hamilton, S.R., Kern, S.E., Preisinger, A.C., Leppert, M., Nakamura, Y., White, R., Smits, A.M., and Bos, J.L. (1988). Genetic alterations during colorectal-tumor development. N. Engl. J. Med., ***319***: 525-532.

Wang, C.Y., Eshleman, J.R., Willson, J.K., and Markowitz, S. (1995). Both transforming growth factor-beta and substrate release are inducers of apoptosis in a human colon adenoma cell line. Cancer Res., ***55***: 5101-5105.

Wang, Q., Lasset, C., Desseigne, F., Frappaz, D., Bergeron, C., Navarro, C., Ruano, E., and Puisieux, A. (1999). Neurofibromatosis and early onset of cancers in hMLH1-deficient children. Cancer Res., ***59***: 294-297.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., and Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. Genes Dev., ***14***: 927-939.

Warthin, A.S. (1913). Hereditary with reference to carcinoma. Arch. Intern. Med., ***12***: 546-555.

Watson, P. and Lynch, H.T. (1993). Extracolonic cancer in hereditary nonpolyposis colorectal cancer. Cancer, ***71***: 677-685.

Weber, J.L. (1990). Informativeness of human (dC-dA)n.(dG-dT)n polymorphisms. Genomics, ***7***: 524-530.

Weber, J.L. and Wong, C. (1993). Mutation of human short tandem repeats. Hum. Mol. Genet., ***2***: 1123-1128.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G., and Lathrop, M. (1992). A second-generation linkage map of the human genome. Nature, ***359***: 794-801.

Wheeler, J.M., Beck, N.E., Kim, H.C., Tomlinson, I.P., Mortensen, N.J., and Bodmer, W.F. (1999). Mechanisms of inactivation of mismatch repair genes in human colorectal cancer cell lines: the predominant role of hMLH1. Proc. Natl. Acad. Sci. U. S. A., ***96***: 10296-10301.

Wheeler, J.M., Loukola, A., Aaltonen, L.A., Mortensen, N.J., and Bodmer, W.F. (2000). The role of hypermethylation of the hMLH1 promoter region in HNPCC versus MSI+ sporadic colorectal cancers. J. Med. Genet., ***37***: 588-592.

Wierdl, M., Dominska, M., and Petes, T.D. (1997). Microsatellite instability in yeast: dependence on the length of the microsatellite. Genetics, ***146***: 769-779.

Willett, W. (1989). The search for the causes of breast and colon cancer. Nature, ***338***: 389-394.

Willett, W.C., Stampfer, M.J., Colditz, G.A., Rosner, B.A., and Speizer, F.E. (1990). Relation of meat, fat, and fiber intake to the risk of colon cancer in a prospective study among women. N. Engl. J. Med., ***323***: 1664-1672.

Williams, C.S., Mann, M., and DuBois, R.N. (1999). The role of cyclooxygenases in inflammation, cancer, and development. Oncogene, ***18***: 7908-7916.

Wrana, J.L., Attisano, L., Wieser, R., Ventura, F., and Massague, J. (1994). Mechanism of activation of the TGF-beta receptor. Nature, ***370***: 341-347.

Wrana, J. and Pawson, T. (1997). Signal transduction. Mad about SMADs. Nature, ***388***: 28-29.

Wu, J., Gu, L., Wang, H., Geacintov, N.E., and Li, G.M. (1999a). Mismatch repair processing of carcinogen-DNA adducts triggers apoptosis. Mol. Cell. Biol., ***19***: 8292-8301.

Wu, Y., Berends, M.J., Mensink, R.G., Kempinga, C., Sijmons, R.H., van Der, Z., Hollema, H., Kleibeuker, J.H., Buys, C.H., and Hofstra, R.M. (1999b). Association of hereditary nonpolyposis colorectal cancer-related tumors displaying low microsatellite instability with MSH6 germline mutations. Am. J. Hum. Genet., ***65***: 1291-1298.

Xu, X., Peng, M., and Fang, Z. (2000). The direction of microsatellite mutations is dependent upon allele length. Nat. Genet., ***24***: 396-399.

Yagi, O.K., Akiyama, Y., Nomizu, T., Iwama, T., Endo, M., and Yuasa, Y. (1998). Proapoptotic gene BAX is frequently mutated in hereditary nonpolyposis colorectal cancers but not in adenomas. Gastroenterology, ***114***: 268-274.

Yamamoto, H., Sawai, H., Weber, T.K., Rodriguez-Bigas, M.A., and Perucho, M. (1998). Somatic frameshift mutations in DNA mismatch repair and proapoptosis genes in hereditary nonpolyposis colorectal cancer. Cancer Res., ***58***: 997-1003.

Yamamoto, H., Perez-Piteira, J., Yoshida, T., Terada, M., Itoh, F., Imai, K., and Perucho, M. (1999). Gastric cancers of the microsatellite mutator phenotype display characteristic genetic and clinical features. Gastroenterology, ***116***: 1348-1357.

Yao, X., Buermeyer, A.B., Narayanan, L., Tran, D., Baker, S.M., Prolla, T.A., Glazer, P.M., Liskay, R.M., and Arnheim, N. (1999). Different mutator phenotypes in Mlh1- versus Pms2-deficient mice. Proc. Natl. Acad. Sci. U. S. A., ***96***: 6850-6855.

Yin, C., Knudson, C.M., Korsmeyer, S.J., and Van Dyke, T. (1997). Bax suppresses tumorigenesis and stimulates apoptosis in vivo. Nature, ***385***: 637-640.

Yokozaki, H. (2000). Distribution of germline BAT-40 poly-adenine tract microsatellite variants in the japanese. Int. J. Mol. Med., ***6***: 445-448.

Zhang, Y., Feng, X., We, R., and Derynck, R. (1996). Receptor-associated Mad homologues synergize as effectors of the TGF-beta response. Nature, ***383***: 168-172.

Zhang, H., Richards, B., Wilson, T., Lloyd, M., Cranston, A., Thorburn, A., Fishel, R., and Meuth, M. (1999). Apoptosis induced by overexpression of hMSH2 or hMLH1. Cancer Res., ***59***: 3021-3027.

Zhang, L., Yu, J., Willson, J.K., Markowitz, S.D., Kinzler, K.W., and Vogelstein, B. (2001). Short Mononucleotide Repeat Sequence Variability in Mismatch Repair-deficient Cancers. Cancer Res., ***61***: 3801-3805.

Zhou, X.P., Hoang, J.M., Cottu, P., Thomas, G., and Hamelin, R. (1997). Allelic profiles of mononucleotide repeat microsatellites in control individuals and in colorectal tumors with and without replication errors. Oncogene, ***15***: 1713-1718.

Appendix A

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
Ibl-1261 B1-D2			Transitions		
	8	A-G	2242	748	ApG
	56	A-G	2267	756	ApA
	83	A-G	2458	820	ApG
	31	C-T	2470	824	CpC
	49	C-T	2626	876	CpG
	12	G-A	3139	1047	GpA
	16	G-A	2782	928	GpC
	29	G-A	2329	777	GpA
	49	G-A	2556	852	GpG
	72	G-A	2863	955	GpA
	101	G-A	3135	1045	GpA
	30	T-C	2980	994	TpT
	85	T-C	2980	994	TpC
Ibl-1261 G1-I2			Transitions		
	117	A-G	3983	1328	ApC
	122	A-G	3976	1326	ApG
	46	C-T	4297	1433	CpC
	49	C-T	3989	1330	CpC
	167	C-T	3987	1329	CpC
	168	C-T	4297	1433	CpC
	18	G-A	4357	1453	GpT
	71	G-A	4525	1509	GpC
	94	G-A	4354	1452	GpA
	99	G-A	4675	1559	GpA
			Transversions		
	162	T-G	4718	1573	TpT

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
Ibl-c5 B1-D2			Transititons		
	1	A-G	2312	771	ApA
	2	A-G	2370	790	ApA
	7	A-G	2565	855	ApC
	7	A-G	2972	991	ApA
	9	A-G	2773	925	ApG
	11	A-G	2272	758	ApA
	11	A-G	2544	848	ApG
	12	A-G	2525	842	ApT
	18	A-G	2435	812	ApC
	33	A-G	2456	819	ApT
	35	A-G	2679	893	ApG
	41	A-G	2647	883	ApC
	17	C-T	3103	1035	CpA
	5	G-A	2658	886	GpA
	9	G-A	2915	972	GpT
	17	G-A	2528	843	GpT
	1	T-C	2961	987	TpG
	4	T-C	2672	891	TpG
	17	T-C	2352	784	TpC
	17	T-C	2382	794	TpC
	17	T-C	2931	977	TpC
	27	T-C	2352	784	TpC
	44	T-C	2781	927	TpG
	46	T-C	2740	914	TpG

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
Ibl-a			Transitions		
B1-D2	65	A-G	2263	755	ApG
	53	A-G	2263	755	ApG
	14	A-G	2267	756	ApA
	46	A-G	2283	761	ApG
	28	A-G	2330	777	ApC
	4	A-G	2379	793	ApA
	56	A-G	2469	823	ApC
	5	A-G	2512	838	ApG
	6	A-G	2540	847	ApA
	27	A-G	2606	869	ApT
	13	A-G	2710	904	ApG
	5	A-G	2750	917	ApT
	26	A-G	2926	976	ApG
	32	A-G	2940	980	ApC
	27	A-G	3118	1040	ApG
	3	C-T	2727	909	CpA
	54	G-A	2768	923	GpA
	50	G-A	2987	996	GpT
	61	T-C	2232	744	TpT
	26	T-C	2235	745	TpG
	6	T-C	2251	751	TpC
	46	T-C	2305	769	TpT
	14	T-C	2491	831	TpT
	37	T-C	2506	836	TpC
	6	T-C	2691	897	TpC
	14	T-C	2751	917	TpG
	42	T-C	2810	937	TpC
	44	T-C	2818	940	TpC
	11	T-C	2844	948	TpA
	24	T-C	2875	959`	TpC
	62	T-C	2878	960	TpC
	23	T-C	2922	974	TpA
	67	T-C	2953	985	TpC
	67	T-C	3048	1016	TpA
	70	T-C	3093	1031	TpT
	58	T-C	3093	1031	TpT
	32	T-C	3112	1038	TpC
	56	T-C	3126	1042	TpC
			Transversions		
	69	T-G	2298	766	TpC
Ibl-a			Transitions		
G1-I2	28	A-G	3799	1260	ApG
	28	A-G	3864	1288	TpG
	33	A-G	4359	1453	TpA
	35	A-G	3917	1306	ApA
	35	A-G	4370	1457	ApA
	44	A-G	3997	1333	ApC
	50	A-G	3927	1309	ApA
	60	A-G	4006	1336	ApG
	68	A-G	4349	1450	ApG
	68	A-G	4370	1457	ApA
	68	A-G	4401	1467	ApC
	75	A-G	4688	1563	ApC
	18	T-C	4586	1529	TpT
	21	T-C	4842	1614	TpG
	23	T-C	4698	1566	TpG
	28	T-C	3865	1289	TpG
	31	T-C	4842	1614	TpG
	44	T-C	4680	1560	TpG
	44	T-C	4842	1614	TpG
	52	T-C	4629	1543	TpA

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
Ibl-1260 B1-D2			Transitions		
	93	A-G	2224	742	ApT
	56	A-G	2242	748	ApG
	39	A-G	2267	756	ApA
	59	A-G	2274	758	ApG
	25	A-G	2280	760	ApG
	59	A-G	2283	761	ApG
	12	A-G	2330	777	ApC
	46	A-G	2368	790	ApG
	63	A-G	2370	790	ApC
	95	A-G	2380	794	ApG
	85	A-G	2420	807	ApT
	9	A-G	2423	808	ApT
	P1-13	A-G	2425	809	ApA
	93	A-G	2458	820	ApC
	59	A-G	2546	849	ApT
	26	A-G	2624	875	ApG
	70	A-G	2702	901	ApG
	15	A-G	2738	913	ApT
	16	A-G	2788	930	ApC
	17	A-G	2905	969	ApG
	50	A-G	2908	970	ApG
	65	A-G	2912	971	ApT
	57	A-G	3008	1003	ApC
	10	A-G	3118	1040	ApG
	44	C-T	2285	762	CpA
	70	C-T	2285	762	CpA
	21	C-T	3009	1003	CpC
	22	G-A	2519	840	GpC
	6	G-A	2871	957	GpA
	55	T-C	2460	820	TpG
	22	T-C	2465	822	TpT
	50	T-C	2500	834	TpC
	93	T-C	2521	841	TpT
	39	T-C	2529	843	TpT
	P1-13	T-C	2535	845	TpT
	19	T-C	2592	864	TpC
	76	T-C	2592	864	TpC
	91	T-C	2619	873	TpT
	63	T-C	2649	883	TpG
	84	T-C	2660	887	TpT
	76	T-C	2690	897	TpT
	16	T-C	2793	931	TpT
	61	T-C	2827	943	TpC
	50	T-C	2850	950	TpT
	74	T-C	2850	950	TpT
	32	T-C	2898	966	TpT
	43	T-C	2958	986	TpT
	95	T-C	2961	987	TpG
	73	T-C	2991	997	TpG
	2	T-C	3018	1006	TpA
	24	T-C	3036	1012	TpC
	2	T-C	3084	1028	TpC
	37	T-C	3093	1031	TpT
	37	T-C	3114	1038	TpG

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
SW480			Transitions		
B1-D2	19	A-G	2273	758	ApA
	20	A-G	2872	958	ApG
	22	A-G	2859	953	ApT
	48	A-G	2338	780	ApG
	51	A-G	2859	953	ApT
	52	A-G	2859	953	ApT
	63	A-G	2452	818	ApA
	81	A-G	2268	756	ApC
	89	A-G	2918	973	ApT
	89	A-G	2928	976	ApG
	93	A-G	2420	807	ApT
	4'	A-G	2415	805	ApC
	8'	A-G	2323	775	ApA
	8'	A-G	2410	804	ApA
	9'	A-G	2706	902	ApG
	10'	A-G	3053	1018	ApT
	17'	A-G	2969	990	ApT
	19'	A-G	2321	774	ApC
	29'	A-G	2755	919	ApG
	34'	A-G	3053	1018	ApT
	40'	A-G	2282	761	ApA
	41'	A-G	2773	925	ApG
	55'	A-G	2250	750	ApT
	58'	A-G	2274	758	ApG
	61'	A-G	2713	905	ApA
	67'	A-G	2644	882	ApC
	73'	A-G	2321	774	ApC
	73'	A-G	2332	778	ApA
	78'	A-G	2514	838	ApG
	82'	A-G	2294	765	ApT
	85'	A-G	2426	809	ApT
	86'	A-G	2324	775	ApT
	96'	A-G	2595	865	ApG
	2	C-T	2795	932	CpA
	44	C-T	2257	753	CpA
	63	C-T	2365	789	CpA
	89	C-T	3032	1011	CpA
	39'	C-T	2254	752	CpT
	41'	C-T	2687	896	CpC
	42'	C-T	2848	950	CpC
	49'	C-T	2955	985	CpT
	51'	C-T	2371	791	CpA
	70'	C-T	2848	950	CpC
	72'	C-T	2799	933	CpA
	16	G-A	2888	963	GpT
	22	G-A	2674	892	GpA
	33	G-A	2357	786	CpG
	37	G-A	2992	998	GpG
	48	G-A	2903	968	GpT
	51	G-A	2673	891	GpG
	62	G-A	3116	1039	GpA
	65	G-A	2662	888	GpC
	78	G-A	2871	957	GpA
	82	G-A	3115	1039	GpG
	20'	G-A	2630	877	GpT
	33'	G-A	2265	755	GpA
	72'	G-A	2422	808	GpA
	53	T-C	3106	1036	TpT
	1'	T-C	2418	806	TpG
	8'	T-C	2765	922	TpT
	29'	T-C	2256	752	TpT
	31'	T-C	2355	785	TpC

32'	T-C	2251	751	TpC
32'	T-C	2809	937	TpT
33'	T-C	2958	986	TpT
41'	T-C	2431	811	TpC
41'	T-C	2875	959	TpC
42'	T-C	2690	897	TpT
48'	T-C	2391	797	TpG
48'	T-C	2787	929	TpA
51'	T-C	2535	845	TpT
62'	T-C	2350	784	TpC
64'	T-C	2698	900	TpC
66'	T-C	2350	784	TpC
70'	T-C	2776	926	TpC
83'	T-C	3036	1012	TpC
92'	T-C	2505	835	TpT
		Transversions		
39'	A-T	2815	939	ApA
		Deletions		
69	T del	2403	801	TpT

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
HCT116			Transitions		
B1-D2	84	A-G	2345	782	ApG
	84	A-G	2366	789	ApG
	59	A-G	2565	855	ApC
	28	A-G	2624	875	ApG
	86	A-G	2652	884	ApG
	3	A-G	2274	758	ApG
	10	A-G	2455	819	ApT
	22	A-G	2924	975	ApA
	33	C-T	2642	881	CpC
	94	C-T	2688	896	CpA
	61	C-T	3025	1009	CpA
	44	C-T	2654	885	CpC
	59	C-T	2783	928	CpC
	22	C-T	2945	982	CpG
	2	G-A	2749	917	GpA
	16	T-C	2350	784	TpC
	1	T-C	2350	784	TpC
	23	T-C	2698	900	TpC
	78	T-C	2818	940	TpC
	76	T-C	2970	990	TpG
HCT116			Transitions		
G1-I2	4	A-G	3852	1284	ApG
	12	A-G	4573	1525	ApT
	17	A-G	3852	1284	ApG
	38	A-G	3794	1265	ApA
	78	A-G	4760	1587	ApG
	87	A-G	4114	1372	ApG
	90	A-G	4598	1533	ApC
	91	A-G	3794	1265	ApA
	4'	A-G	4309	1437	ApG
	27'	A-G	4222	1408	ApG
	41'	A-G	4071	1357	ApG
	63'	A-G	4317	1439	ApC
	68'	A-G	4071	1357	ApG
	69'	A-G	4049	1350	ApC
	92'	A-G	4299	1433	ApC
	92'	A-G	4486	1496	ApG
	83	C-T	4526	1509	CpT
	41'	C-T	3875	1292	CpG
	68'	C-T	3875	1292	CpG
	2	G-A	3895	1299	GpC
	16	G-A	4391	1464	GpA
	90	G-A	4791	1597	GpA
	69'	G-A	4221	1407	GpA
	4	T-C	4295	1432	TpG
	13	T-C	4177	1393	TpC
	20	T-C	4266	1422	TpT
	21	T-C	4469	1490	TpA
	29	T-C	4586	1529	TpT
	37	T-C	4095	1365	TpT
	59	T-C	4266	1422	TpT
	59	T-C	4581	1527	TpC
	87	T-C	4873	1625	TpT
	38'	T-C	4266	1422	TpG
	47'	T-C	3825	1275	TpT
	56'	T-C	4854	1618	TpC
	63'	T-C	3905	1302	TpG
	78'	T-C	3841	1281	TpC
	79'	T-C	4673	1558	TpT
	79'	T-C	4698	1566	TpT
			Deletions		
	80'	Adel	4093	1365	ApG
			Transversions		
	92'	A-T	4606	1536	ApA
	79'	T-A	4691	1564	TpA

Cell line, *APC* region	Clone, ID	Mutation	Nucleotide position	Codon	Target sequence
HCT116			Transitions		
+chr3	74	A-G	2312	771	ApA
B1-D2	36	A-G	2330	777	ApC
	7	A-G	2433	811	ApG
	21	A-G	2647	883	ApC
	41	A-G	2708	903	ApC
	90	A-G	2746	916	ApC
	28	A-G	2746	916	ApC
	40	A-G	2940	980	ApC
	60	A-G	3132	1044	ApC
	61	A-G	2564	855	ApA
	23	C-T	2255	742	TpT
	40	C-T	2278	760	CpT
	59	C-T	2786		CpA
	34	G-A	2449	817	GpG
	58	G-A	2450	817	GpC
	11	G-A	2528	843	GpT
	83	G-A	2971	991	GpA
	34	T-C	2505	835	TpT
	73	T-C	2536	846	TpC
	60	T-C	3087	1029	TpA
			Insertions		
	14	G insertion	2359	787	ins
HCT116			Transitions		
+chr3	20	A-G	4662	1544	ApG
G1-I2	57	A-G	4616	1539	TpC
	85	A-G	3879	1293	ApC
	78	A-G	4045	1349	ApG
	83	A-G	4646	1549	ApC
	2'	A-G	4349	1450	ApG
	7'	A-G	4000	1334	ApA
	8'	A-G	4374	1458	ApC
	9'	A-G	3929	1310	ApG
	11'	A-G	3961	1321	ApG
	16'	A-G	3852	1284	ApG
	17'	A-G	3861	1287	ApG
	33'	A-G	3797	1266	ApT
	84	C-T	4044	1348	CpA
	8	G-A	3889	1297	GpA
	34	G-A	4414	1472	GpC
	64	G-A	4426	1476	GpC
	87	G-A	3974	1325	GpC
	26	T-C	4320	1440	TpC
	41	T-C	4177	1393	TpC
	58	T-C	3801	1267	TpC
	59	T-C	4101	1367	TpC
	70	T-C	4177	1393	ApG
	7'	T-C	4197	1399	TpC
	8'	T-C	3980	1327	TpG
	8'	T-C	4177	1393	TpC
	8'	T-C	4320	1440	TpC
	8'	T-C	4562	1521	TpG
	9'	T-C	4148	1383	TpC
	11'	T-C	4329	1443	TpC
	16'	T-C	4332	1444	TpC
	18'	T-C	3791	1264	CpA
	18'	T-C	4167	1389	TpT
	19'	T-C	4020	1340	TpT
	19'	T-C	4817	1606	TpG
	27'	T-C	4122	1374	TpG
			Insertions		
	17'	Gins	3876	1292	GpA
			Deletions		
	47'	Adel	3823	1275	ApG

Human Molecular Genetics, 2000, Vol. 9, No. 18 2707–2713

Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells

Andrea L. Bacon, Susan M. Farrington and Malcolm G. Dunlop*

University of Edinburgh Department of Oncology and MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received 24 July 2000; Revised and Accepted 13 September 2000

Determinants of instability at a given microsatellite repeat merits investigation in view of relevance to understanding evolution of mutations at such sequences in human populations. The microsatellite *D2S123* was studied as a paradigm CA repeat marker. Furthermore, this marker is one of a recommended panel used in molecular screening for hereditary non-polyposis colorectal cancer (HNPCC). In this investigation we show that the mutation rate at the *D2S123* locus is markedly influenced by intra-allelic sequence variation within the repetitive tract itself. We employed a novel approach to characterize the nature of instability at *D2S123*, by utilizing cells derived from a non-tumour lineage, which harbour a dominant negative mismatch repair (MMR) mutation and a mutator phenotype. Individual alleles were typed using a semi-quantitative small pool PCR technique and this demonstrated substantial allele-these specific bias in susceptibility to mutation at the *D2S123* locus. In support of these *in vitro* data, bias in allele mutation rate was also observed in tumours from 41 HNPCC patients, which was dependent on constitutional genotype. Sequencing of cell line and patient DNAs revealed that short alleles are significantly more susceptible to mutation due to the presence of uninterrupted CA repeats. Long *D2S123* alleles are intrinsically more stable because of a TA interspersion within the repetitive tract. In addition to extending understanding of mutation at CA repeat dinucleotide tracts, these findings have considerable relevance both to screening programmes and to correlation of microsatellite instability (MSI) with colon cancer survival. The manifestation of tumour MSI may be substantially influenced by constitutional genotype.

INTRODUCTION

Microsatellites occur ubiquitously throughout the genome and mutations within these simple repetitive tracts are frequent as evidenced by heterozygosity rates (1). In studies of poly(CA/GT) repeats in humans, high rates of mutation per locus per gamete per generation (5.6×10^{-4}) have been revealed (2). The inherently unstable nature of microsatellites results in frequent alterations in the length of the repeat tracts making many of them highly polymorphic (1,3). Differences in mutation rate are evident at alternative microsatellite loci (1) and sequence variation can significantly affect individual microsatellite stability. Variations in microsatellite flanking sequences (4,5) and in repeat length (6,7) have been demonstrated to contribute to heterogeneity in mutation rate. In addition, correlation of instability with the degree of perfection of a repeat has been documented. In *Escherichia coli* and yeast, presence of an interspersion within a dinucleotide tract results in its stabilization (5,8); this is also observed in the presence of defective mismatch repair (MMR) (8). Locus-by-locus analyses reveal that interspersed repeat markers are also relatively more stable in the fruit fly *Drosophila* (9), in the germline of human populations (10) and in tumours from cancer patients (11). Repeat sequence interruption is also important at a clinical level. Expansion of triplet repeats gives rise to the human trinucleotide repeat disorders such as fragile X, Huntington's disease and spinocerebellar ataxia type 1 (SCA1) (12,13). Stability of both *SCA1* and *FMR1* (fragile X gene) alleles are conferred by interruption of the contiguous repeat (14–18).

The high degree of polymorphism makes microsatellites invaluable tools for use in genetic mapping, in DNA forensic studies, as population markers and in a variety of other applications. Mechanisms affecting mutation rate within a given locus have been less extensively characterized, although there are studies documenting such variation. For example, in a population study of meioses in pentameric and tetrameric markers, it was shown that different alleles at single loci varied in their rate of mutation (10). However, investigation of mechanisms affecting microsatellite stability in family studies requires very large numbers of meiosis to be screened in order to identify mutations (2). This places severe practical limitations on such studies as well as on investigations using sperm DNA (19).

To investigate factors influencing stability at given microsatellite alleles we employed a novel approach using a 'sensitized' system. We utilized a lymphoblast cell line from a patient who developed colorectal cancer. The cell line is defective in MMR and displays a mutator phenotype but is derived from a non-tumourous tissue. There is evidence that mechanisms generating mutations in microsatellite unstable (MSI+) tumours share similarities with those resulting in the evolution

*To whom correspondence should be addressed. Tel: +44 131 467 8439; Fax: +44 131 343 2620; Email: malcolm.dunlop@hgu.mrc.ac.uk

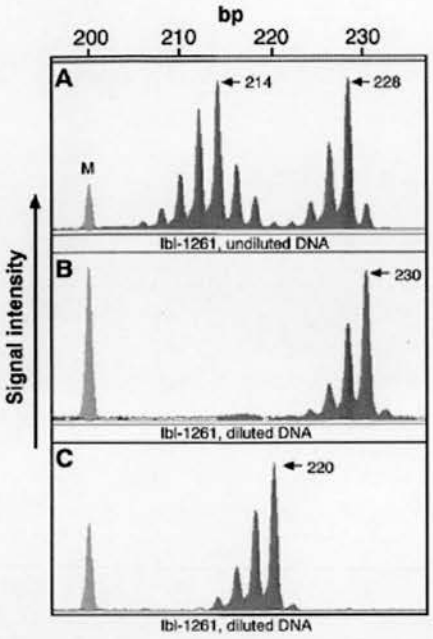


Figure 1. Genotyping of individual *D2S123* alleles in cell line lbl-1261. (**A**) Constitutional genotype of DNA from lbl-1261 determined from the preparation of undiluted DNA. A short and a long allele of 214 and 228 bp, respectively, are clearly detected. (**B** and **C**) Individual mutant alleles of 230 and 220 bp detected in lbl-1261 by SP-PCR. M, marker peak at 200 bp.

of such sequences in the germline (11,20) and a mutator phenotype might exacerbate the process of germline evolution at repetitive markers. Hence, study of determinants of marker stability in MSI$^+$ MMR-defective cells has relevance to the understanding of the process of evolution of repeat sequences in human populations (11,20).

Defects in mismatch repair result in the accelerated accumulation of frameshift mutations within microsatellite sequences (21–23) occurring predominantly at mononucleotide and dinucleotide repeat tracts. Over 90% of hereditary non-polyposis colorectal cancer (HNPCC) cases display microsatellite instability (MSI) (24,25), usually as a result of germline defects in one of at least five MMR genes (*hMLH1*, *hMSH2*, *hPMS2*, *hPMS1* and *hMSH6*) (26,27). During tumour development, the wild-type allele is inactivated by loss of heterozygosity (LOH), mutation or epigenetic silencing causing complete loss of MMR activity. Some MSI$^+$ sporadic colorectal tumours are also defective in MMR by nature of somatic mutations, LOH or by epigenetic silencing due to hypermethylation of the *hMLH1* promoter (28). Once inactivated, the consequent failure to repair DNA replication errors that arise due to DNA polymerase slippage manifests as MSI.

In a cell line defective in MMR due to a dominant negative *hPMS2* mutation, we show that alleles at a microsatellite repeat locus are differentially stabilized and demonstrate the sequence determinant of this mutation bias. The CA repeat microsatellite studied here is one of a panel of five markers recommended for use in MSI analysis in colorectal cancer (29,30). In addition to having general importance in understanding the mechanisms that lead to instability at repetitive sequences, this work also has specific relevance to the influence of patient genotype on the manifestation of tumour MSI and also on correlation of MSI with clinicopathological features.

RESULTS

Determination of allele bias in MMR-deficient cells

Genotyping of *D2S123* alleles from MMR-defective lymphoblast cell line lbl-1261 revealed alleles of 228 and 214 bp (Fig. 1). In all, a total of 270 and 115 alleles were typed by small pool PCR (SP-PCR) for lbl-1261 and a control lymphoblast cell line (lbl-c5), respectively. Ninety alleles (33.3%) from lbl-1261 templates exhibited length variation compared with only two alleles (1.7%) of the control cell line ($\chi^2 = 42.3$, $P < 7 \times 10^{-11}$) (Figs 1 and 2, Table 1). The small number of mutants found in lbl-c5 may be due to PCR error or indeed could be genuine mutations. However, their small number relative to those in lbl-1261 indicates that PCR artefacts do not interfere significantly with the detection of mutants in the dilute DNA. These data are in accordance with a previous study of mutation frequency in this cell line (31). However, we noted a substantial bias in the alleles from which mutations were derived. In all, 72 of 90 mutants (80%) in lbl-1261 were clustered around the shorter (214 bp) allele (Fig. 2A), indicating a substantial bias for mutation at that locus ($\chi^2 = 42.3$, $P < 0.00001$). Mutant alleles were assumed to be derived from the progenitor allele closest in size in the undiluted DNA, since studies of microsatellite mutations in human pedigrees, human cell lines and artificial constructs in yeast have shown that most microsatellite mutations involve only one or two repeat units (2,8,32).

To ensure that the effect was not due to misassignment of mutant alleles, we re-analysed the data using a threshold to exclude reductions in the 228 bp allele. We assumed that only mutants of <216 bp (a single repeat expansion of the short progenitor allele) were derived from the shorter progenitor. This confirmed the highly significant mutation bias at the shorter allele ($\chi^2 = 8.45$, $P < 0.0037$). These observations are not due to PCR bias at the smaller allele since equal numbers of short and long alleles were detected in lbl-1261. Furthermore, insertion and deletion mutants of both progenitor alleles were observed (Fig. 2).

These observations indicate a substantial and statistically significant allele-specific bias in mutation rate. We were interested to determine the nature of this bias, especially since it conflicts with the expectation that MSI correlates with increased repeat length and number of repeating units (7,33)

Thirty-nine Scottish individuals were genotyped at the *D2S123* locus. A total of eight different alleles were identified (210–230 bp) consistent with the CEPH data (http://www.cephb.fr/cephdb/) (Table 2). Alleles were noticed to cluster in two distinct size groups, long (~228 bp) or short (~214 bp) and the frequency of alleles in each of these two groups were almost identical between our cohort and the CEPH data.

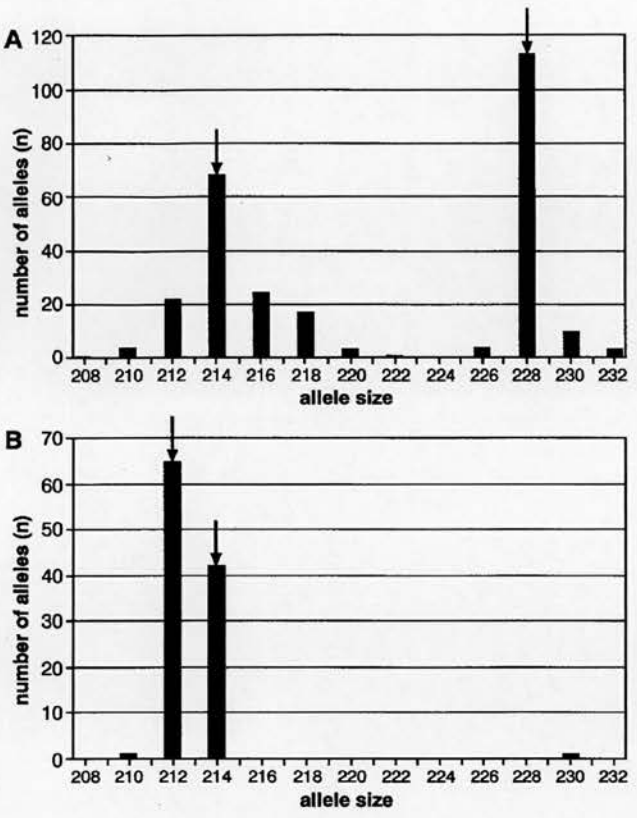


Figure 2. Distribution of progenitor and mutant *D2S123* alleles in MMR-deficient (lbl-1261) (**A**) and -proficient (lbl-c5) (**B**) lymphoblast cell lines as detected by SP-PCR. Arrows indicate wild-type alleles as identified by analysis of undiluted DNA. Equal numbers of short and long alleles and their respectively assigned mutants were identified indicating that there is not a PCR preference in detecting shorter alleles. A significant mutation bias at the 214 bp progenitor allele in lbl-1261 is demonstrated ($\chi^2 = 42.3$, $P < 0.00001$).

Table 1. Summary of mutant alleles detected by SP-PCR in cell lines lbl-1261 and lbl-c5

Cell line	Total no. of alleles	Mutants (frequency)
lbl-1261	270	90 (0.33)
lbl-c5	115	2 (0.17)

lbl-1261 is significantly more unstable than control cell line lbl-c5 ($\chi^2 = 42.3$, $P < 7 \times 10^{-11}$).

Mutation rate within the *D2S123* CA repeat is predicted by sequence content

The sequence for *D2S123* logged in GenBank is $(CA)_{13}(TA)(CA)_{15}$ (GenBank accession no. Z16551). Individual alleles were sequenced to determine whether sequence differences between long and short alleles might explain the substantial mutation bias. Initially undiluted lbl-1261 DNA was gel separated and individual alleles sequenced independently. This demonstrated that the 228 bp allele comprised of a $(CA)_{28}$ repeat tract split into two runs of $(CA)_{13}$ and $(CA)_{15}$ by a TA dinucleotide, concurring with the sequence logged in GenBank. However, the shorter 214b bp allele comprised of an uninterrupted $(CA)_{22}$ repeat with no interspersing TA dinucleotide (Fig. 3). Genotyping and sequencing of 15 *D2S123* alleles from template DNA of a cohort of Scottish individuals showed that various length alleles classified as long always contained the TA interspersion, whereas various lengths of short alleles were invariably uninterrupted poly(CA) repeats. This analysis indicated that alleles clustered around 214 bp (short) contain perfect uninterrupted CA repeats. In contrast long alleles consistently contain a TA interspersion within the CA tract.

In order to confirm definitively that mutant alleles arose predominantly from uninterrupted alleles in cell line lbl-1261, *D2S123* alleles were PCR amplified from undiluted lbl-1261 template DNA and cloned. Individual alleles were sized and sequenced. Analysis of 28 clones confirmed that every long allele analysed contained the TA interruption and every short allele contained perfect CA repeats, thereby confirming that the presence of a TA interspersion confers stability on the CA repeats. Mutation was due to length variation of the number of core CA repeats in every case, implicating replication slippage as the causative mechanism of instability at this locus (Table 3). Cell line lbl-c5 was heterozygous for two short alleles and PCR cloning of undiluted DNA revealed two short uninterrupted alleles of $(CA)_{21}$ and $(CA)_{22}$ repeats. As expected in this MMR-proficient cell line, no mutations were detected in the clones despite both alleles having no interspersion of the repeat tract (data not shown).

Taken together these data indicate that long alleles (~228 bp) invariably contain a TA interspersion in the repeat whereas short alleles (~214 bp) contain uninterrupted $(CA)_n$ repeats that are inherently more unstable. This instability is unmasked in MMR-deficient cell lines.

Allele-specific bias of mutations at *D2S123* in DNA MMR-deficient tumours

To determine whether allele bias might also influence MSI phenotype in MMR-deficient tumours, we genotyped *D2S123* alleles in matched normal and tumour DNA samples from 41 colorectal cancer patients with MSI$^+$ tumours. These have been characterized previously and exhibit MSI at four or more markers. In some cases the germline mutation has been identified (unpublished data) (34). Normal tissue was genotyped for *D2S123* alleles and the stability of each allele assessed in matched tumour DNA. Thirty of the 41 tumours exhibited mutation at one or more *D2S123* alleles and overall 37 of the 82 alleles (45%) had mutations. There was a significant bias in the frequency of mutations at the short alleles compared with long alleles (Fig. 4). Of 53 small wild-type alleles identified, 32 (60%) displayed instability in the tumour tissue compared with only 5 of 29 (17%) large alleles ($\chi^2_1 = 12.4$, $P = 0.00043$). Thus, constitutional allele sequence is a determinant of the propensity for instability at a given locus in the presence of defective MMR. This has important implications for the classification of clinical material with respect to MSI status.

Table 2. Comparison of *D2S123* allele frequencies between CEPH database and 39 Scottish individuals

Number	Size of fragment[a] (bp)	Frequency (CEPH data)	Frequency (Scottish data)	Category
1	228	0.304	0.244	Long
2	212	0.339	0.41	Short
3	216	0.107	0.089	Short
4	214	0.125	0.128	Short
5	215	0.054	0.026	Short
6	198	0.018	0.00	Short
7	226	0.018	0.064	Long
8	210	0.036	0.013	Short
u	230	0.00	0.026	Long

Our genotyping data for a small Scottish cohort is in agreement with those from CEPH family analysis. We did not identify any 198 bp alleles; however, a previously undocumented allele size of 230 bp (u) was observed.
[a]Our allele sizes (shown) genotyped exactly 1 bp longer than the corresponding CEPH alleles. This is likely to be due to difference in primer length.

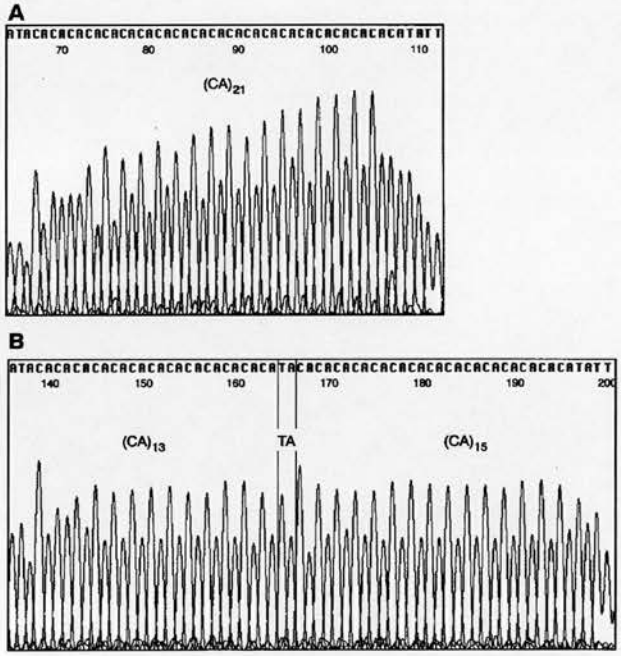


Figure 3. Sequence analysis of the two progenitor *D2S123* alleles in lbl-1261. The short allele (**A**) consists of an uninterrupted $(CA)_{22}$ tract. The long wild-type allele (**B**) consists of $(CA)_{13}$ repeats followed by a TA dinucleotide and then another run of $(CA)_{15}$ repeats.

DISCUSSION

We studied a paradigm CA repeat microsatellite locus, *D2S123*, and have shown that a cell line derived from non-tumour tissue displays a mutator phenotype and that the resultant accumulation of mutations is predicted by host genotype at each allele. *D2S123* is used routinely in tumour MSI analysis and shows inter-allelic sequence variability for an interruption within the repeat tract. By genotyping and direct sequencing of individual alleles, we have demonstrated that defective MMR results in insertion and deletions of CA repeat units, which accounts for the variation in size of mutant alleles. In addition, alleles without a TA interspersion within the repeat tract are significantly more susceptible to instability as a consequence of defective MMR. Equal numbers of long and short alleles and their respectively assigned mutants were identified. Furthermore, insertion and deletion mutants derived from both progenitor alleles were detected. Such findings argue strongly against the trivial explanation that these observations are due to PCR bias in detecting shorter mutants at this locus. The analysis of a cohort of MSI$^+$ colorectal cancer patients provides further confirmation that patient genotype directly influences manifestation of a mutator phenotype and the likelihood of MSI status being scored accurately.

Other studies that have addressed differences between loci in human populations and in colorectal cancers also show that perfect repeats are more susceptible to instability (8,9,11). However, this study is the first to demonstrate that variant interruptions can occur between alleles at the same marker loci and that these have a marked effect on individual allele stability. We show that the stabilizing effects of a variant repeat is apparent even in the presence of defective MMR, supporting a previous report (8). These results suggest that different alleles at the same marker loci can display genotypic variation, substantially affecting susceptibility to mutation.

Several studies have suggested that the mechanism of microsatellite mutation in MSI$^+$ tumours shares similarity with the evolution of such sequences in the genome (11,20). *D2S123* population allele frequencies from both our own data and that of the CEPH database highlights higher frequencies of short alleles. The CEPH database (http://www.cephb.fr/cephdb/) documents six short *D2S123* alleles compared with two long ones (Table 2). This would be expected if short alleles with perfect repeats are more susceptible to mutation as our results from MMR-defective cells suggest. These mutations could then be fixed as new alleles. The absence of many length vari-

Table 3. Sizing and sequencing of *D2S123* clones from lbl-1261

Allele size[a] (bp)	Sequence	No. of clones	Category
208	$(CA)_1$	3	Short
210	$(CA)_2$	6	Short
212	$(CA)_2$	4	Short
214	$(CA)_2$	5	Short
216	$(CA)_2$	2	Short
224	$(CA)_{12}TA(CA)_1$	1	Long
226	$(CA)_{12}TA(CA)_1$	1	Long
228	$(CA)_{13}TA(CA)_1$	5	Long
230	$(CA)_{14}TA(CA)_1$	1	Long

The sequences of the CA repeat region only are indicated.
[a]Size of alleles as shown by ABI 310 analysis. Allele sizes of 208–216 nucleotides are classed as short and those of 224–230 nucleotides are long (see text). All short alleles contain perfect CA repeats in contrast to long alleles that have a TA interspersion. Variation in allele size is wholly accounted for by differences in the number of CA repeats.

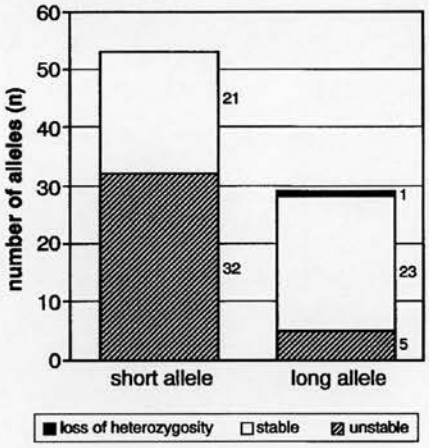


Figure 4. Tumour DNA instability at *D2S123* alleles in 41 MSI+ colorectal cancer patients. Long and short alleles were compared between matched normal and tumour DNA and assessed for stability. There is a significant excess of mutations at short alleles ($\chi^2 = 12.4$, $P = 0.00043$).

ants of the long alleles is also consistent with our sequence data showing that long alleles invariably contain the stabilizing interspersion.

It is generally accepted that replication slippage is the major mechanism causing new mutations in microsatellites (35). The results presented here are consistent with models in which a variant interruption causes stabilization by encouraging the perfect realignment of the two strands following their dissociation during such DNA polymerase slippage. However, the possibility that interruption of the repeat tract may alter some unusual structure associated with the repeats and subsequently reduce the rate of slippage cannot be wholly discounted (8).

The majority of mutations occurring at CA microsatellites have been shown to involve small length changes in which one or two repeats are altered (8,10,36). Mutations of repeat tracts in cancer genes of MSI+ colorectal tumours characteristically display small frameshift mutations (37,38). The mutations identified here at the *D2S123* locus also involve the gain or loss of one or two repeat units in the majority of cases. In the cell line lbl-1261, short mutant alleles always contained perfect repeats, whereas long mutant alleles consistently possessed a TA interruption. The absence of any short mutants with a repeat interruption, or long mutants without an interruption or of mutants with duplicated TA interspersions argues against the occurrence of large sequence alterations in excess of a few repeats. This is consistent with hPMS2 being involved in the repair of small insertion/deletion loops (39).

Defects in MMR result in genome-wide accumulation of mutations at repetitive microsatellite sequences. We have used a cell line derived from a non-tumor lineage with a dominant negative MMR mutation to investigate inter-allelic MSI. This system has distinct advantages over using cancer cell lines. Cancer cell lines are subject to selection pressures for mutations that provide the neoplastic cell with a distinct growth advantage, thus evolving clonally and making them largely homogenous for mutations at any given microsatellite. In addition they accumulate multiple mutations and abnormalities, making the dissection of events resulting directly and exclusively from MMR defects difficult.

These data show that host genotype at CA repeat sequences can influence the ability of available marker sets to assign MSI status to any individual tumour. Therefore, these findings have important clinical relevance regarding MSI screening strategies and the effects of individual patient genotype on these analyses. Tumour MSI status is used to determine whether HNPCC may be discounted or whether analysis of MMR genes is required. Additionally there have been a number of reports indicating that MSI status may be used as a predictor of survival and can be employed as a clinical tool with which to give patient survival estimates (40–41). An accurate assessment of MSI status is therefore of critical importance. The other four microsatellite markers in the recommended panel (29) are entered in GenBank as uninterrupted $(CA)_n$ or $(A)_n$ repeats. However, the CEPH database (http://www.cephb.fr/cephdb/) indicates that they are polymorphic for different sized alleles. Such analysis by length may mask further underlying individual genotype differences at these markers. In any case, the phenomenon described here for *D2S123* alone is of importance and the influence of patient genotype at this particular marker may be particularly critical in the diagnosis of borderline MSI cases. The data from this study highlight the need for a well-characterized set of diagnostic markers in which allelic variance and the intrinsic effect on stability is well understood. Suggestions that one or a few microsatellite markers are sufficient to assess MSI status should perhaps be treated with caution, especially when there may be significant implications for both the patient and the family (29,30,42). The marker studied here is logged in GenBank as an interspersed microsatellite. However we have shown that this does not reflect the underlying complexity of this locus. Many other markers may share similar intrinsic variances in allelic stability and the phenomenon may be widespread.

MATERIALS AND METHODS

Lymphoblast cell lines and tumour samples

Epstein–Barr virus (EBV) transformed lymphoblast cell lines were cultured from a healthy control individual (lbl-c5) and from the non-tumour tissue of a patient with colorectal cancer who had a family history of the disease (lbl-1261). Cell line lbl-1261 is derived from patient 6 referred by Parsons *et al.* (31) and displays a mutator phenotype due to a dominant negative mutation in hPMS2 (31,43).

Matched tumour and normal DNAs were previously isolated from 41 colorectal cancer patients shown to have defective MMR and a mutator phenotype. The causative MMR gene mutations have been defined in a number of cases (unpublished data) (34).

All patients with suitable material available whose tumours fulfilled MSI criteria were analysed in this study, regardless of whether or not *D2S123* showed length variation in tumour DNA. There is potential bias because MSI criteria include *D2S123* genotyping and so some patients with homozygous long alleles may be underscored with respect to MSI status. However, this does not impact adversely on the findings of these studies.

SP-PCR

We devised an assay to genotype *D2S123* alleles, using an SP-PCR strategy in order to investigate the mutational behaviour of individual alleles. A similar approach has been used previously to detect mutations within populations of wild-type alleles (44,45). DNA from EBV transformed cell lines was diluted to a final concentration of 15–20 pg per PCR reaction to give up to three input molecules of each *D2S123* allele per amplification (assuming 6 pg of DNA per diploid genome). Due to the dilute nature of the DNA, alleles were detected in ~30% of analyses. Fluorescently labelled *D2S123* primers were used in all PCR reactions (31). PCR amplifications were performed in a final volume of 25 µl. Final reaction concentrations were 1× PCR Buffer II (Boehringer, Mannheim, Germany), 0.2 mM dNTPs, 3.7 pM oligonucleotide primer and 0.87 U of Expand high fidelity PCR system enzyme mix. Reactions were prepared in 96 well plates. DNA-free controls were prepared in 16 of the wells in each plate and positive controls containing 100 ng of cell line DNA were prepared in 2 wells in every plate. Amplification was performed using an Omnigene PCR system thermal cycler (Hybaid) at 94°C for 3 min for 1 cycle, 94°C for 1 min, 55°C for 1 min, 72°C for 1 min for 35 cycles, 72°C for 5 min for 1 cycle. Two microlitres of each PCR reaction, including positive and negative controls, were analysed on an ABI 310 Automated Genetic Analyser, using Genescan software. For each DNA sample 100–250 SP-PCR products were generated and analysed. The frequency of mutant alleles in each cell line was expressed as the number of alleles which were mutant in length divided by the total number of alleles detected (normal and mutant). Accordingly, percentages are not exact contents of cells with alterations, but relative values of alleles. Differences between MSI frequency in the two cell lines was evaluated by χ^2 test and significance taken at 5%. Mutant allele origin was determined by assignment to the progenitor allele closest in size according to previous studies and our own experiments (2,8,32). Differences in mutation frequency of each progenitor allele was evaluated by χ^2 test. To account for the possibility that some mutants did not derive from the progenitor allele closest in size, a further more stringent assessment was performed by assuming that only mutants of <216 bp were derived from the smaller allele.

Sequencing *D2S123* alleles

D2S123 alleles were amplified from undiluted cell line DNA using non-fluorescent primers and the aforementioned high fidelity PCR system and electrophoresed on a 3% Nusieve GTG agarose (Bioproducts, Rockland, ME) 1% 'Hi Pure' low EEO agarose (Biogene, Kimbolton, UK) gel in 1× TBE buffer at 30 V overnight and individual progenitor alleles gel purified (Qiagen, Crawley, UK). Sequencing was performed using PRISM Ready Big Dye Terminator Cycle Sequencing kit with AmpliTaq DNA polymerase, FS (Taq-FS; Perkin Elmer/Applied Biosystems, Branchburg, NJ) and Applied Biosystems DNA sequencer model 373A or 377, according to the manufacturer's instructions. *D2S123* alleles from lbl-1261 and lbl-c5 were cloned into the TA cloning vectors (Invitrogen, Groningen, The Netherlands) according to the manufacturer's instructions. Genotyping and cycle sequencing of transformants was performed as above.

Allele-specific bias in MSI tumours

One hundred nanograms of normal and tumour DNA templates were used in 50 µl PCR reactions as described. Analysis of individual allele shifts was made by comparison of Genescan profiles on the ABI analyser from normal and tumour DNA. Whenever there were doubts about the veracity of a mutation we did not include it. This applied to cases where the patient harboured two short alleles. On occasions it was questionable as to whether one or both alleles had mutated. Such cases were scored as shifting at just one of the alleles. Mutation frequency may therefore be marginally underestimated if both alleles had mutated and the presence of any wild-type allele in the tumour was due to contamination from surrounding normal mucosa. Observed differences in mutation frequency at each allele were evaluated by a χ^2 analysis.

ACKNOWLEDGEMENTS

We are grateful to Bert Vogelstein for cell line lbl-1261. We acknowledge expert technical assistance from Sheila McBeath. This work was supported by the following grants: Cancer Research Campaign (SP2326/0201), Scottish Health Department (K/MRS/50/C2723), Melville Trust, Edinburgh University Cancer Research Endowment Fund and Urquhart Charitable Trust. A.L.B. was funded by an MRC PhD Student Fellowship, S.M.F. by an RSE Personal Research Fellowship and M.G.D. by an MRC Clinician Scientist Fellowship.

REFERENCES

1. Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., Millasseau, P., Vaysseix, G. and Lathrop, M. (1992) A second-generation linkage map of the human genome. *Nature*, **359**, 794–801.
2. Weber, J.L. and Wong, C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**, 1123–1128.
3. Weber, J.L. (1990) Informativeness of human $(dC\text{-}dA)_n$.$(dG\text{-}dT)_n$ polymorphisms. *Genomics*, **7**, 524–530.

4. Andreassen, R., Egeland, T. and Olaisen, B. (1996) Mutation rate in the hypervariable VNTR g3 (D7S22) is affected by allele length and a flanking DNA sequence polymorphism near the repeat array. *Am. J. Hum. Genet.*, **59**, 360–367.
5. Bichara, M., Schumacher, S. and Fuchs, R.P. (1995) Genetic instability within monotonous runs of CpG sequences in *Escherichia coli*. *Genetics*, **140**, 897–907.
6. Xu, X., Peng, M. and Fang, Z. (2000) The direction of microsatellite mutations is dependent upon allele length. *Nature Genet.*, **24**, 396–399.
7. Wierdl, M., Dominska, M. and Petes, T.D. (1997) Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics*, **146**, 769–779.
8. Petes, T.D., Greenwell, P.W. and Dominska, M. (1997) Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics*, **146**, 491–498.
9. Goldstein, D.B. and Clark, A.G. (1995) Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Res.*, **23**, 3882–3886.
10. Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J. and Rolf, B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.*, **62**, 1408–1415.
11. Sturzeneker, R., Bevilacqua, R.A., Haddad, L.A., Simpson, A.J. and Pena, S.D. (2000) Microsatellite instability in tumors as a model to study the process of microsatellite mutations. *Hum. Mol. Genet.*, **9**, 347–352.
12. Ashley, C.T.J. and Warren, S.T. (1995) Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.*, **29**, 703–728.
13. Gordenin, D.A., Kunkel, T.A. and Resnick, M.A. (1997) Repeat expansion—all in a flap? *Nature Genet.*, **16**, 116–118.
14. Chung, M.Y., Ranum, L.P., Duvick, L.A., Servadio, A., Zoghbi, H.Y. and Orr, H.T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genet.*, **5**, 254–258.
15. Eichler, E.E., Holden, J.J., Popovich, B.W., Reiss, A.L., Snow, K., Thibodeau, S.N., Richards, C.S., Ward, P.A. and Nelson, D.L. (1994) Length of uninterrupted CGG repeats determines instability in the FMR1 gene. *Nature Genet.*, **8**, 88–94.
16. Kunst, C.B., Leeflang, E.P., Iber, J.C., Arnheim, N. and Warren, S.T. (1997) The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. *J. Med. Genet.*, **34**, 627–631.
17. Hirst, M.C., Grewal, P.K. and Davies, K.E. (1994) Precursor arrays for triplet repeat expansion at the fragile X locus. *Hum. Mol. Genet.*, **3**, 1553–1560.
18. Snow, K., Tester, D.J., Kruckeberg, K.E., Schaid, D.J. and Thibodeau, S.N. (1994) Sequence analysis of the fragile X trinucleotide repeat: implications for the origin of the fragile X mutation. *Hum. Mol. Genet.*, **3**, 1543–1551.
19. Jeffreys, A.J., Neumann, R. and Wilson, V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell*, **60**, 473–485.
20. Di Rienzo, A., Donnelly, P., Toomajian, C., Sisk, B., Hill, A., Petzl-Erler, M.L., Haines, G.K. and Barch, D.H. (1998) Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. *Genetics*, **148**, 1269–1284.
21. Ionov, Y., Peinado, M.A., Malkhosyan, S., Shibata, D. and Perucho, M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
22. Thibodeau, S.N., Bren, G. and Schaid, D. (1993) Microsatellite instability in cancer of the proximal colon. *Science*, **260**, 816–819.
23. Strand, M., Prolla, T.A., Liskay, R.M. and Petes, T.D. (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature*, **365**, 274–276.
24. Aaltonen, L.A., Peltomaki, P., Leach, F.S., Sistonen, P., Pylkkanen, L., Mecklin, J.P., Jarvinen, H., Powell, S.M., Jen, J. and Hamilton, S.R. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science*, **260**, 812–816.
25. Aaltonen, L.A., Salovaara, R., Kristo, P., Canzian, F., Hemminki, A., Peltomaki, P., Chadwick, R.B., Kaariainen, H., Eskelinen, M., Jarvinen, H. *et al.* (1998) Incidence of hereditary nonpolyposis colorectal cancer and the feasibility of molecular screening for the disease. *N. Engl. J. Med.*, **338**, 1481–1487.
26. Peltomaki, P. and Vasen, H.F. (1997) Mutations predisposing to hereditary nonpolyposis colorectal cancer: database and results of a collaborative study. The International Collaborative Group on Hereditary Nonpolyposis Colorectal Cancer. *Gastroenterology*, **113**, 1146–1158.
27. Miyaki, M., Konishi, M., Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Yasuno, M., Igari, T., Koike, M., Chiba, M. and Mori, T. (1997) Germline mutation of MSH6 as the cause of hereditary nonpolyposis colorectal cancer. *Nature Genet.*, **17**, 271–272.
28. Herman, J.G., Umar, A., Polyak, K., Graff, J.R., Ahuja, N., Issa, J.P., Markowitz, S., Willson, J.K., Hamilton, S.R., Kinzler, K.W. *et al.* (1998) Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc. Natl Acad. Sci. USA*, **95**, 6870–6875.
29. Boland, C.R., Thibodeau, S.N., Hamilton, S.R., Sidransky, D., Eshleman, J.R., Burt, R.W., Meltzer, S.J., Rodriguez-Bigas, M.A., Fodde, R., Ranzani, G.N. *et al.* (1998) A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.*, **58**, 5248–5257.
30. Perucho, M. (1999) A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.*, **59**, 249–256.
31. Parsons, R., Li, G.M., Longley, M., Modrich, P., Liu, B., Berk, T., Hamilton, S.R., Kinzler, K.W. and Vogelstein, B. (1995) Mismatch repair deficiency in phenotypically normal human cells. *Science*, **268**, 738–740.
32. Hoff-Olsen, P., Meling, G.I. and Olaisen, B. (1998) Variation in mutation rate and direction between tetranucleotide STR loci in human colorectal carcinomas. *Ann. Hum. Genet.*, **62**, 1–7.
33. Sia, E.A., Kokoska, R.J., Dominska, M., Greenwell, P. and Petes, T.D. (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol. Cell Biol.*, **17**, 2851–2858.
34. Farrington, S.M., Lin-Goerke, J., Ling, J., Wang, Y., Burczak, J.D., Robbins, D.J. and Dunlop, M.G. (1998) Systematic analysis of hMSH2 and hMLH1 in young colon cancer patients and controls. *Am. J. Hum. Genet.*, **63**, 749–759.
35. Levinson, G. and Gutman, G.A. (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.*, **15**, 5323–5338.
36. Henderson, S.T. and Petes, T.D. (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol. Cell Biol.*, **12**, 2749–2757.
37. Rampino, N., Yamamoto, H., Ionov, Y., Li, Y., Sawai, H., Reed, J.C. and Perucho, M. (1997) Somatic frameshift mutations in the BAX gene in colon cancers of the microsatellite mutator phenotype. *Science*, **275**, 967–969.
38. Parsons, R., Myeroff, L.L., Liu, B., Willson, J.K., Markowitz, S.D., Kinzler, K.W. and Vogelstein, B. (1995) Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.*, **55**, 5548–5550.
39. Kolodner, R.D. and Marsischky, G.T. (1999) Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.*, **9**, 89–96.
40. Gryfe, R., Kim, H., Hsieh, E.T., Aronson, M.D., Holowaty, E.J., Bull, S.B., Redston, M. and Gallinger, S. (2000) Tumor microsatellite instability and clinical outcome in young patients with colorectal cancer. *N. Engl. J. Med.*, **342**, 69–77.
41. Lothe, R.A., Peltomaki, P., Meling, G.I., Aaltonen, L.A., Nystrom-Lahti, M., Pylkkanen, L., Heimdal, K., Andersen, T.I., Moller, P. and Rognum, T.O. (1993) Genomic instability in colorectal cancer: relationship to clinicopathological variables and family history. *Cancer Res.*, **53**, 5849–5852.
42. Frazier, M.L., Sinicrope, F.A., Amos, C.I., Cleary, K.R., Lynch, P.M., Levin, B. and Luthra, R. (1999) Loci for efficient detection of microsatellite instability in hereditary non-polyposis colorectal cancer *Oncolo. Rep.*, **6**, 497–505.
43. Nicolaides, N.C., Littman, S.J., Modrich, P., Kinzler, K.W. and Vogelstein, B. (1998) A naturally occurring hPMS2 mutation can confer a dominant negative mutator phenotype. *Mol. Cell Biol.*, **18**, 1635–1641.
44. Jeffreys, A.J., Tamaki, K., MacLeod, A., Monckton, D.G., Neil, D.L. and Armour, J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.*, **6**, 136–145.
45. Monckton, D.G., Wong, L.J., Ashizawa, T. and Caskey, C.T. (1995) Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum. Mol. Genet.*, **4**, 1–8.

Nucleic Acids Research, 2001, Vol. 29, No. 21 4405–4413

Hypermutability at a poly(A/T) tract in the human germline

Andrea L. Bacon, Malcolm G. Dunlop and Susan M. Farrington*

University of Edinburgh Department of Oncology and MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK

Received July 3, 2001; Revised and Accepted September 6, 2001

ABSTRACT

Poly(A/T) tracts are abundant simple sequence repeats (SSRs) within the human genome. They constitute part of the coding sequence of a variety of genes, encoding polylysine stretches that are important for protein function. Assessment of poly(A/T) tract stability is also used to identify microsatellite unstable colorectal cancers, which are characteristic of tumours defective in DNA mismatch repair. Despite their importance, little is known about the stability of poly(A/T) SSRs in the human germline. We have determined the stability of a paradigm poly(A/T) tract, BAT-40, by study of population allele frequencies, mutation frequency in families and mutation frequency in sperm DNA. We show that the locus is polymorphic, with a level of heterozygosity of 59.7%. Germline mutation was observed in 13 of 187 germline transmissions (7.0%) in 10 families suggesting BAT-40 is unstable in the germline. Further evidence for germline instability at BAT-40 was provided by small pool PCR analysis of matched blood and sperm DNA templates, revealing a significantly elevated frequency of mutation in the germline ($P < 0.001$). These findings provide insight into poly(A/T) tract stability in the germline. They also have relevance to the study of gene expression and to determination of microsatellite instability in tumours.

INTRODUCTION

Simple sequence repeats (SSRs) occur ubiquitously throughout the genome. Many are highly polymorphic, making them of particular importance to the study of evolution and the mapping of disease genes (1). SSRs such as poly(A/T) and $(CA)_n$ are routinely used in determining microsatellite instability (MSI) status in mismatch repair (MMR) deficient colorectal cancers (CRCs) (2–5). The proportion of markers that display mutational shifts in the tumour directs subsequent analysis for germline MMR mutations that lead to this MSI phenotype (6). Factors affecting the mutation frequency of microsatellites have been studied in MMR deficient tumours since analysis is rapid and material readily available (7,8). We have previously demonstrated that constitutional genotype at a given microsatellite locus influences the propensity for instability in the presence of defective MMR (7). Such studies highlight the need for a well-characterised panel of markers to be used for such assessments, in order for them to be employed with confidence.

It has been shown that mechanisms generating mutations in microsatellite unstable (MSI^+) tumours have relevance to understanding the evolution of such sequences in the germline (8). Investigation of determinants of germline mutation at SSR loci is laborious and frequently requires analysis of many hundreds of gametes in family studies (9,10). However, the development of small pool PCR (SP-PCR) techniques in studying germline stability at minisatellites has facilitated investigations of such mutations at other SSRs (11,12).

To date, many studies of germline mutation at SSR loci have focused on understanding sequence instability of the trinucleotide repeat disorders (13–16). In addition, investigations have been carried out on germline stability of dinucleotide repeat markers, including long CA stretches (10). However, there has been little investigation of the stability of mononucleotide tracts in the germline. This is surprising since poly(A/T) repeats are the most abundant simple repetitive sequence motif in the human genome (17) largely due to the poly(A/T) tails of scattered retrotransposed sequences such as long (LINEs) and short interspersed elements (SINEs) (17,18). Coding poly(A/T) sequence tracts have been identified with repeat lengths of up to 27 bp and within introns they may occur up to 70 bp in repeat length (17). Any process influencing the fidelity of replication at coding sequence mononucleotide tracts will clearly have important functional effects. The transforming growth factor beta receptor type 2 (*TGFBR2*) gene contains a $poly(A/T)_{10}$ tract in exon 3 that has been shown to be mutated in up to 90% of MSI^+ tumours, resulting in inactivation of the gene (19–21).

In view of the prevalence of poly(A/T) stretches and their functional relevance we were interested to gain insight into the inherent stability of such sequences. The microsatellite BAT-40 is a paradigm mononucleotide marker consisting of 40 adenine repeats located in intron 2 of the 3-beta-hydroxysteroid dehydrogenase (*3-beta-HSD*) gene on chromosome 11 (22). BAT-40 is highly sensitive to the effects of defective MMR, since it is susceptible to mutation in >95% MSI^+ tumours and thus is used routinely in the analysis of MSI (5,21,23,24). Previous studies have also demonstrated that BAT-40 exhibits significant

*To whom correspondence should be addressed at: Colon Cancer Genetics Group, MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK. Tel: +44 131 467 8422; Fax: +44 131 343 2620; Email: susan.farrington@hgu.mrc.ac.uk

polymorphism within populations (25,26). Hence we hypothesised that BAT-40 and other long poly(A/T) repeats might be unstable in the germline.

We analysed germline stability at the BAT-40 locus as a paradigm poly(A/T) tract, in order to gain insight into the generation of new mutations at such sequences. Assessment of the degree of mutability at such a locus might have considerable relevance to the generation of mutations at that locus in MMR deficient tumours.

MATERIALS AND METHODS

DNA sample groups

Genotyping was carried out on the constitutional DNA of 102 unrelated Scottish individuals and 35 unrelated CEPH family members from the nine families listed below.

A Scottish Family, K-435, was used for pedigree analysis. Family relationships were confirmed by genotyping, using a panel of microsatellite markers (data not shown). This family was identified previously as being an HNPCC kindred with affected individuals displaying tumour instability as determined using a panel of microsatellite markers. Proband MD-473 was previously determined to be heterozygous at BAT-40 with two alleles differing by 12 bp in length thus making individual allele identification obvious. MD numbers represent our laboratory sample identification system. DNA samples were available from the peripheral blood leukocytes of 20 individuals from this family.

DNA from nine CEPH families (66, 1331, 1341, 1346, 1362, 1377, 1423, 13293 and 13294) was used for further pedigree analysis to provide a total of 176 germline transmissions for study.

SP-PCR was carried out on matched constitutional and germline DNA samples from two further unrelated individuals. MD-949 carries a germline mutation in the MMR gene, human *MLH1*, resulting in a deletion of exon 12 (codons 347–470). MD-c1 is a healthy control individual.

Preparation of constitutional DNA and BAT-40 genotyping

Constitutional DNA was extracted from blood using Nucleon II DNA extraction protocol (Scotlab Bioscience, Strathclyde).

BAT-40 alleles were amplified from DNA templates in triplicate using primers described previously (24). These primers amplify a 126 bp product containing the standard 40 A residues according to the genomic sequence of the *3-beta-HSD* gene (GenBank accession no. M38180) (22). We were not able to assess the number of adenine repeats in a given sized allele directly, since repeated attempts at sequencing across the BAT-40 locus were unsuccessful. Therefore repeat length is based on the theoretical predicted amplified sequence both in this and in the majority of other studies that report BAT-40 repeat length.

PCR reactions were performed in a final volume of 25 µl using the Expand High Fidelity PCR system (Boehringer Mannheim, Germany). Final reaction concentrations were 1× PCR buffer II, 0.2 mM dNTPs, 100 ng oligonucleotide primer, 100 ng DNA and 0.87 U of Expand high fidelity PCR mix. Amplifications were performed using an Omnigene PCR thermal Cycler (Hybaid) at 94°C for 3 min for 1 cycle; 94°C for 1 min, 55°C for 1 min, 72°C for 1 min for 35 cycles; 72°C for 5 min for 1 cycle. PCR products were size analysed using an ABI310 Automated Genetic Analyser, using Genescan software.

Sperm DNA preparation and SP-PCR

DNA was extracted from sperm as described by Jeffreys *et al.* (27). Pelleted semen was rinsed three times with 20 ml 1× SSC followed by six washes with 20 ml 1× SSC and 1% SDS to lyse any seminal leukocytes and epithelial cells. The residual sperm pellet was incubated in 1× SSC and 1 M 2-mercaptoethanol at room temperature for 5 min and the reduced sperm lysed by addition of SDS to 1%. Sperm DNA was collected after phenol extraction by ethanol precipitation and resuspended in Tris–EDTA pH 7.7. DNA concentration was calculated using a spectrophotometer to measure the optical density of DNA samples in triplicate and also by running samples against standards of known DNA quantity using gel electrophoresis.

SP-PCR and analysis of PCR products was performed as described (7). DNA from matched blood and sperm samples were serially diluted to a final concentration of 15–20 pg per PCR reaction. This results in an estimated five to six template BAT-40 alleles in total, per reaction (assuming 6 pg of DNA per diploid genome). Limiting dilutions were carried out and final DNA concentrations resulted in the detection of a product in ~30% of analyses corresponding to products that represent single DNA templates (11,28). PCR reactions were then performed as described above in 96 well plates using the Expand High Fidelity PCR System (Boehringer Mannheim, Germany). Avoidance of contamination was paramount when amplifying dilute DNA templates. Therefore all reactions were carried out in a Class-2 containment hood. All pipettes and plastics used in the preparation of the SP-PCR reactions were UV irradiated for 20 min in a Template Tamer (Oncor). Buffer solution and sterile water were opened under sterile conditions and also subjected to UV irradiation. On each plate, 16 wells were DNA free to provide negative controls and positive controls containing 100 ng of undiluted sample DNA was prepared in two wells on every plate to ensure reproducibility of ABI310 profiles between plates. Matched sperm and blood DNA samples and SP-PCRs were prepared simultaneously, using the same reagents to allow direct comparison. Amplifications were performed as above and SP-PCR products and positive controls were size analysed using ABI310 Automated Genetic Analyser and Genescan software. All 16 negative controls from every SP-PCR plate were also analysed and if a product was observed in any negative sample, the entire plate was discarded.

Determination of the origin of new alleles

In instances where the origin of 'new/mutant' alleles was inferred, this was done in the same manner as in published studies (e.g. 10). The origin of 'new/mutant' alleles is such that if there were two possibilities, the shortest mutational step was considered to be the actual one. For example, if one progenitor allele differed by one repeat and the other by two repeats, when compared to the mutant, a one step mutation was inferred. If two progenitor alleles exhibited the same difference when compared to the new/mutant one, the origin was declared ambiguous.

Statistical analysis

Statistical comparison of population allele size frequency at the BAT-40 locus was carried out using a Mann–Whitney U test on the Minitab (V.13) statistical package. Significance was taken at the 5% level.

For comparison of mutation levels between matched sperm and blood samples, the frequency of mutant alleles in each sample was expressed as the number of alleles that were mutant in length divided by the total number of alleles detected (normal and mutant). Accordingly the frequency of mutants was not the exact number of cells with alterations but represents the relative proportions of alleles. Statistical analyses were then performed using a chi-squared analysis on Minitab (V.13) statistical package, and significance taken at the 5% level.

RESULTS

The BAT-40 poly(A/T) locus is polymorphic

BAT-40 genotypes were defined in 104 unrelated Scottish individuals and 35 unrelated CEPH family members (Table 1). Representative ABI310 profiles are shown in Figure 1. PCR products displaying a single complex of peaks with a near normal distribution were counted as homozygous (Fig. 1A). Those with extra peaks were regarded as BAT-40 heterozygous (Fig. 1B–D). The allele traces of the BAT-40 mononucleotide marker are complex with 'stutter' peaks evident due to DNA polymerase slippage. However, 'bona fide' allele sizes are taken to be the predominant peak in each separate peak complex in accordance with previous studies (Fig. 1) (26,29). The predominant peak is that with the greatest peak area as indicated by the ABI310 genetic analyser software. Our own previous analysis of this locus by SP-PCR analysis of multiple single alleles in three individuals has also validated this method of allele sizing at the BAT-40 locus. The most predominant peaks as genotyped from constitutional DNA are detectable as individual alleles by SP-PCR (Fig. 2). PCR error is evident when amplifying BAT-40, by nature of the stutter bands that are observed (Figs 1 and 2). However, reproducibility of the prominent peaks in a given individual assures confidence in the sizes given (see Materials and Methods). Where the genotype of an individual could not be confirmed by reproducibly detecting the same peaks in the allele trace, these individuals were discarded from further analysis (three cases).

The distribution of BAT-40 heterozygous genotypes also indicates that amplification and detection of two BAT-40 alleles of different sizes is due to the difference in genotype and not technical artefact (Table 1).

Allele frequency and distribution for both the Scottish and the CEPH cohorts are shown in Figure 3. Of the 139 samples analysed, a total of 83 demonstrated heterozygosity at the BAT-40 locus (59.7%). Levels of heterozygosity were similar between the cohorts, 58.7% (61/104) in the Scottish population and 62.9% (22/35) in the CEPH cohort. As expected for a highly polymorphic marker the overall distribution of allele sizes was not significantly different between these two populations ($P = 0.056$) (Fig. 3). Allelic size variation was from –15 to +11 as compared to the most frequent allele. Taking into account variation in the cohorts studied, these data are in line with a previous study reporting polymorphism at this

Table 1. Levels of heterozygosity at the BAT-40 locus

	Allele set (bp)	Frequency detected in Scottish cohort (% $n = 104$)	Frequency detected in CEPH cohort (% $n = 35$)
Heterozygotes	108/119	0	2.9
	108/123	1.9	0
	109/124	0	2.9
	108/125	1	0
	109/122	1	0
	111/123	0	2.9
	111/124	1	0
	112/124	1	0
	118/119	1	0
	118/121	2.9	0
	118/123	1.9	0
	118/124	1	0
	119/120	1	0
	119/121	1.9	0
	119/122	6.7	5.7
	119/123	1	2.9
	119/124	1.9	0
	119/126	1	0
	119/127	1	2.9
	120/121	1	0
	120/122	5.7	11.4
	120/123	9.6	5.7
	120/124	1.9	5.7
	120/125	1.0	0
	121/122	1.0	0
	121/123	3.8	8.6
	121/124	0	2.9
	122/123	0	2.9
	122/124	1.9	0
	122/125	1.0	0
	123/124	1.0	0
	123/125	1.9	0
	123/128	1	0
	124/125	0	2.9
	124/127	0	2.9
	124/134	1	0
Homozygotes	117/117	1.9	0
	118/118	2.9	0
	119/119	4.8	0
	120/120	3.8	8.6
	121/121	3.8	0
	122/122	4.8	0
	123/123	16.3	22.9
	124/124	3.8	5.7

Frequency of individual BAT-40 genotypes in the Scottish and CEPH cohorts analysed are indicated. Allele sets are given in base pairs and are grouped according to whether or not the genotype is heterozygous.

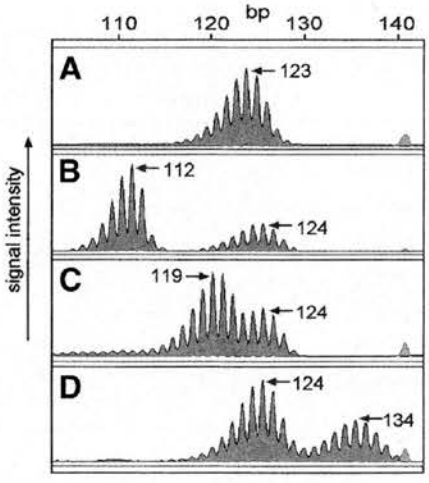


Figure 1. ABI310 traces of BAT-40 poly(A/T) PCR products show the polymorphic nature of this microsatellite marker in samples from a Scottish cohort. Blood DNA shows a single complex of peaks, the highest being 123 bp (**A**). (**B–D**) Heterozygosity at BAT-40 as illustrated by separate peak complexes.

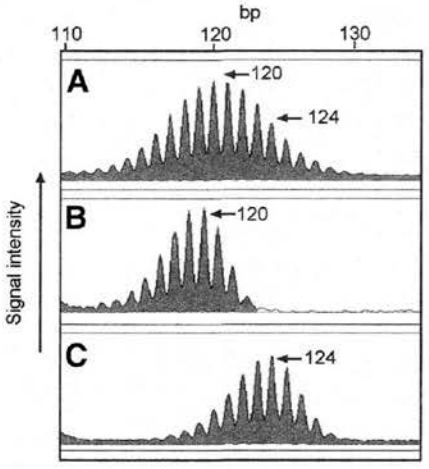


Figure 2. Genotyping of the BAT-40 locus in constitutional DNA is validated by SP-PCR analysis. (**A**) Constitutional BAT-40 allele sizes in cell line lbl-1261 are revealed as 120 and 124 bp from analysis of undiluted DNA. (**B** and **C**) In SP-PCR analysis of DNA from the same cell line, individual alleles of 120 and 124 bp are easily determined and confirm the genotype revealed from the undiluted DNA. (Data taken from ref. 33).

mononucleotide marker for a large number of different alleles, and are suggestive of the frequent generation of new alleles at this locus (25). The most frequent allele in both cohorts in this study (123 bp) corresponds to a 37-adenine tract as calculated from the genomic sequence (GenBank accession no. M38180). However, the BAT-40 heterozygosity reported here differs in both frequency and size distribution to that observed in a Japanese study despite the use of similar methodology (29).

Germline hypermutability at BAT-40 in pedigree analysis

Since BAT-40 displays high levels of polymorphism in populations and has been previously demonstrated to be extremely

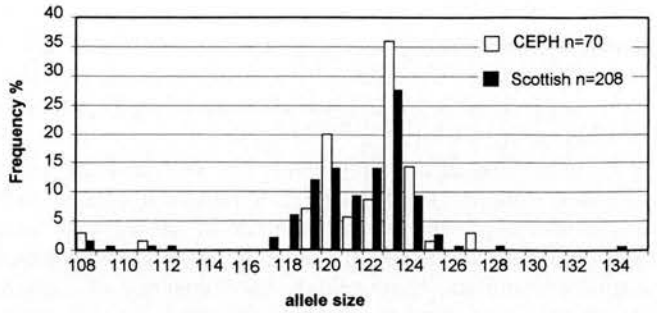


Figure 3. Comparisons of BAT-40 allele frequencies between Scottish and CEPH populations. The sizes of each allele are given in base pairs. The estimated size of the standard BAT-40 allele with 40 adenine repeats is 126 bp as calculated from the predicted PCR product size (GenBank accession no. M38180). There is no statistical difference in the distribution of alleles within the two cohorts ($P = 0.073$).

susceptible to instability in MMR deficient tumours (21), we reasoned that BAT-40 may be inherently unstable and that germline mutations might be detectable in family studies. A Scottish family, K-435, was chosen to determine whether a high level of germline mutation occurs within this population. Proband MD-473 had previously been identified as heterozygous at BAT-40, with two distinct sized alleles at this locus (112/124). Analysing the meiotic stability of BAT-40 alleles that are easily distinguished by size allows for the most accurate assessment of individual allele stability at a complex locus such as BAT-40. DNA from 20 available individuals from K-435 was genotyped at the BAT-40 locus (Fig. 4A). There were 11 germline transmissions available for study. Within the family there was striking evidence of a germline mutation in the allele transmission from MD-1303 to MD-449. MD-1303 is heterozygous for BAT-40 with an allele set of 120/124 but her daughter (MD-449) is homozygous for two 112 BAT-40 alleles (Fig. 4). DNA was unavailable from the father of MD-449 who is very likely to have carried at least one 112 allele, inferred from sibling and progeny genotypes. Therefore the mutation is implicated as being maternal in origin showing loss of repeats at the BAT-40 locus. The 112/112 homozygous allele is highly unlikely to have arisen by dropout of the larger 120 or 124 bp allele during PCR because four family members including MD-439, the sister of MD-449, had 112/124 genotypes that were easily detected under the PCR conditions used (Fig. 4). This indicates that the technique reliably detects the larger alleles. Furthermore, the 120 bp allele in MD-1303 was faithfully amplified (Fig. 4B). Genotyping for all members of this family was confirmed in triplicate and previous genotyping at microsatellite markers confirmed that MD-1301 was indeed the mother of the twins MD-449 and MD-439 (data not shown).

The observation of a germline mutation in only 11 allele transmissions is striking, since only one mutation event in 3000–5000 transmissions has been reported for CA repeats (10). This led us to further investigate the possibility that BAT-40 is highly unstable in the germline and that mutant alleles might be transmitted in subsequent generations, to manifest as polymorphism within the population.

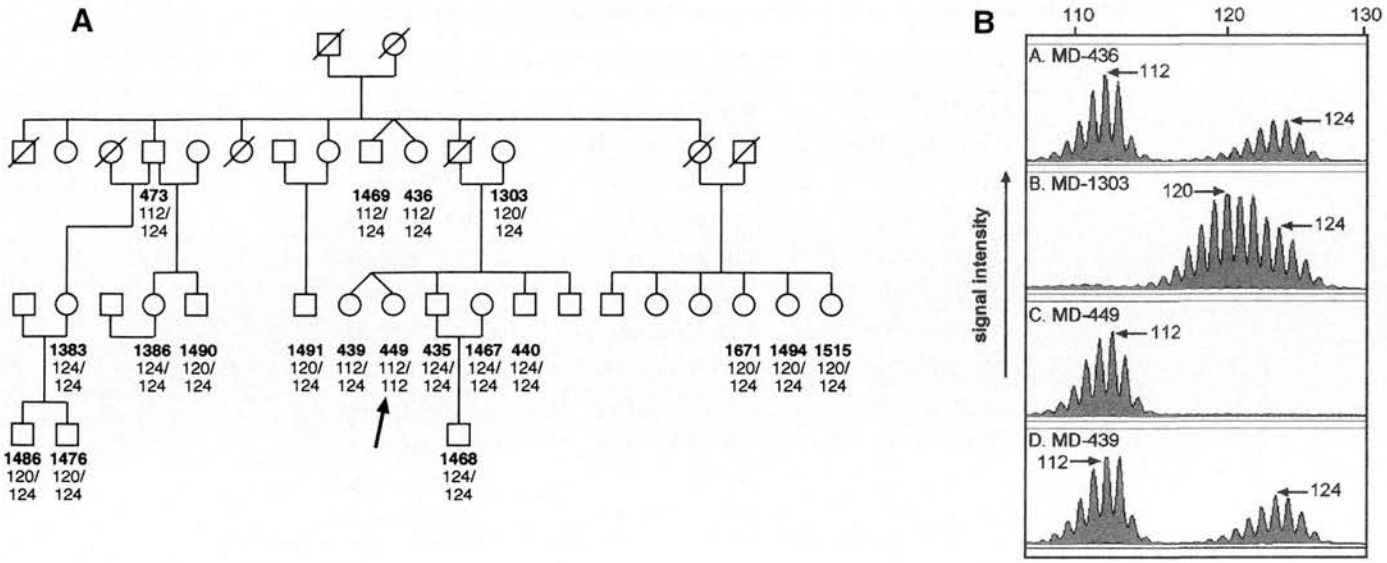


Figure 4. (**A**) BAT-40 genotypes of blood DNA from available samples of pedigree K-435. A single incidence of transmissible germline hypermutability is highlighted. Mother MD-1303 has allele set 120/124 whereas daughter MD-449 (arrow) is homozygous for 112. (**B**) The ABI310 profiles of the BAT-40 alleles of MD-1303 and MD-449 are shown. There is no indication of the presence of either of MD-1303's alleles in MD-449. Although DNA was not available from the father it is likely that he carried at least one 112 allele as inferred from siblings such as his sister MD-436. The profile of MD-439 also demonstrates that there is not a problem in detecting the 124 allele in the presence of the 112 allele.

Germline hypermutability in CEPH family analysis

To further explore germline instability at BAT-40, we analysed BAT-40 alleles in a CEPH family panel. Nine CEPH families were genotyped at the BAT-40 locus, totalling 176 germline transmissions analysed, and 12 putative mutations (6.8%) were identified (Table 1). However, in all cases, heterozygous parental alleles differed by only a few base pairs and the mutations indicated involved small (1 bp) changes (Table 1 and Fig. 5). Of the 88 maternal transmissions analysed, three were mutant at BAT-40 (3.4%) and of the 88 paternal transmissions analysed, nine were mutant (10.2%). This difference was not statistically significant ($\chi^2 = 3.22$, $P = 0.073$). Insertions and deletions appeared to occur equally.

The mutations identified in CEPH families provided further support for our initial observation that BAT-40 is hypermutable in the germline. The CEPH data add further weight to the identification of the germline mutation in family K-435 and also to the evidence from the high levels of heterozygosity at BAT-40 in the caucasian population study demonstrating high levels of heterozygosity at BAT-40. However, although the CEPH mutations were reproducible, the small changes observed in the complex BAT-40 profile led us to devise a further, rigorous method to analyse susceptibility of this locus to mutation in the germline.

Inherent hypermutablity of BAT-40 in the germline demonstrated by small-pool PCR analysis of sperm DNA

SP-PCR analysis of germline DNA has important advantages over family studies for analysing germline stability at complex loci (11,15). The method overcomes the practical constraints encountered during pedigree analyses, which suffer limitations from the small number of mutants that can be identified. In contrast, many hundreds of gametes can be analysed from a single semen sample and consequently, a greater variation in allele size changes may be available for identification. In addition, the dilution of the DNA sample aids unambiguous allele identification at a hypermutable locus. Mutation frequency as detected in sperm DNA has been shown to reflect estimations from studies in pedigrees (11). Comparisons of sperm DNA and constitutional DNA templates have shown that SP-PCR reliably discriminates alleles in both and that there are no demonstrable differences in technical artefact between the two sample templates (11). In addition SP-PCR has been demonstrated to reliably detect differences in intra-allelic mutation frequency in sperm DNA (11,14–16). For SP-PCR of sperm DNA, study subjects were selected as being constitutionally heterozygous at the BAT-40 locus with individual wild-type alleles easily distinguished by size. MD-c1 had allele sizes 120/124 and MD-949 had alleles of size 121/124. Correct identification of constitutional allele sizes was further supported in the SP-PCR analysis where individual alleles of the same predominant allele size were detected (Fig. 6). Approximately 100 SP-PCR products were analysed per sample. BAT-40 allele sizes typed from constitutional and sperm DNA templates by SP-PCR are shown in Figure 7. Mutant alleles were detected in sperm DNA by comparison with constitutional genotype (Fig. 6). The frequency of mutant alleles detected in each sample is summarised in Table 3. In both MD-c1 and MD-949 matched samples, there was a significantly greater number of mutant alleles detected in sperm DNA compared to that of matched blood leukocyte DNA samples ($\chi^2 = 19.32$, $P < 0.001$ and $\chi^2 = 13.82$, $P < 0.001$ for MD-c1 and MD-949, respectively). A total of 9/198 (4.5%) alleles in the leukocyte DNA were mutant compared to a total of 64/255 (25.1%) mutant alleles in the sperm templates, indicating an almost 6-fold increase in mutation accumulation in the germline. The proportion of mutant alleles in blood and sperm was

Table 2. Putative BAT-40 mutations detected in germline transmissions of CEPH families

Family	Genotype (father)	Genotype (mother)	Genotype (child)
66	122/123 (f, -01)	124/127 (m, -02)*	123/126 (c/m, -03)
	123/123 (fm, -12)*	124/127, (mm, -14)	124/127 (m, -02)
1331	123/123 (f, -01)*	119/123 (m, -02)	119/124 (c/m, -17)
1341	120/123 (f, -01)*	123/125 (m, -02)[a]	119/125 (c/f, -05)
	120/123 (f, -01)*	123/125 (m, -02)[a]	119/125 (c/f, -08)
1346	123/123 (f, -01)*	122/127 (m, -02)	124/127 (c/f, -08)
	123/123 (f, -01)*	122/127 (m, -02)	124/127 (c/f, -09)
1362	120/120 (fm, -15)	121/123 (mm, -16)*	120/120 (m, -02)
	121/123 (f, -01)*	120/120 (m, -02)	120/120 (c/f, -04)
1377	120/124 (ff, -10)	120/122 (mf, -11)*	120/121 (f, -01)
	120/121 (f, -01)*	119/122 (m, -02)	122/123 (c/m, -08)
13293	120/123 (f, -01)*	108/109 (M, -02)	109/124 (c/m, -09)

Transmissions in which a germline mutation was indicated are described. The genotype of the child along with both parents is presented. The inferred origin is highlighted by an asterisk. The CEPH pedigree number, individual identification number and family relationship is given as standard CEPH nomenclature.
[a]In one case the sample failed to amplify and genotype was inferred from the relatives.

equivalent in MD-c1 and MD-949 samples suggesting that constitutional heterozygous DNA MMR gene mutation does not influence mutation rate in the male gamete in this case.

Although both addition and deletion mutations were identified, we observed a bias towards repeat losses over gains in the sperm ($\chi^2 = 11.0$, $P < 0.001$ and $\chi^2 = 10.3$, $P < 0.001$ for MD-c1 and MD-949 sperm samples, respectively).

DISCUSSION

Long poly(A/T) tracts are abundant in the human genome, occurring at the 3′ untranslated region (UTR) of genes (30) and within intronic sequences as well as coding regions (17). Length variation at such repetitive sequences has importance with respect to gene expression and function (21,30–32). In addition, poly(A/T) markers play an important role in the classification of microsatellite unstable CRCs (6). Here we report a series of studies aimed at defining germline stability of a paradigm poly(A/T) repeat locus. By analysis of two cohorts we show that BAT-40 is a highly polymorphic locus with an observed level of heterozygosity of 59.7%. This is similar to a previous analysis of a CEPH cohort in which a level of heterozygosity of 72% was reported. However, the level of BAT-40 heterozygosity detected here and in a CEPH cohort by Zhou *et al.* (25) is considerably higher that that observed in a Japanese cohort (14.6%) (29). Although this might be explained in part by variation in allele heterozygosity between populations, the Japanese study cohort were from hospital based samples and may not be representative of the true Japanese population frequencies. Our confirmation that BAT-40 is a highly polymorphic locus suggests that generation of new alleles by slippage and mutation at this locus might be quite common.

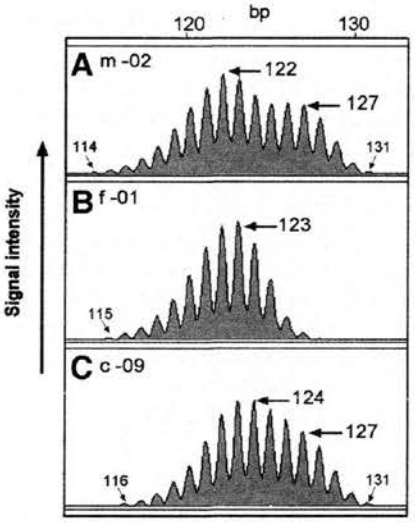


Figure 5. Representative example of a putative BAT-40 germline mutation in CEPH family 1346. While the mother -02 (**A**) has a BAT-40 genotype of 122/127, the father -01 (**B**) appears homozygous for a 123 bp allele. The mother's 127 bp allele is detected in child -09 (**C**) but the most predominant peak in the first complex is at 124 bp. This would indicate a 1 bp mutation at BAT-40 had occurred in the germline of the father. Sizing of the extreme stutter bands also indicates a 1 bp mutation of the father's 123 bp allele in c-09 and confirms the presence of the 127 allele derived from m-02.

Instability at BAT-40 is well documented in MMR deficient tumours (6,21) and we have also observed instability at this locus in cells derived from normal tissue that are deficient in MMR (33). These data also indicate that this repeat locus might be particularly unstable. Identification of a germline mutation at BAT-40 in a Scottish pedigree suggests that this locus is also

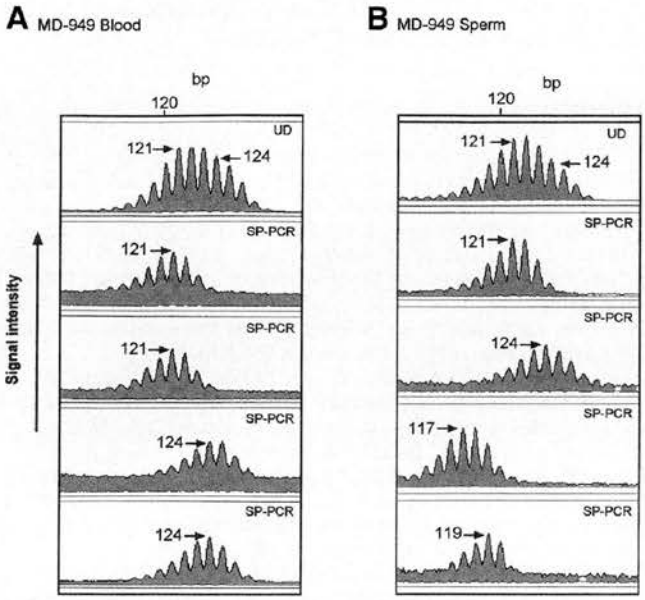


Figure 6. Representative ABI310 traces of BAT-40 alleles detected by SP-PCR in matched blood (**A**) and sperm (**B**) DNA. Almost all BAT-40 SP-PCR products amplified from blood DNA revealed individual alleles with predominant peaks of the same size as those in undiluted (UD) DNA. For MD-949 these were 121 and 124 bp. The majority of BAT-40 SP-PCR products amplified from sperm DNA were also of wild-type allele size as shown. However, a significant number of mutant alleles were detected. Mutants of 117 and 119 bp are illustrated here.

highly unstable in the germline. Explanations such as failure to amplify larger alleles as an explanation for apparent germline mutations, such as that in MD-449, are highly unlikely since larger alleles were detected reliably in the presence of a 112 allele in other family members. Our initial observation in a single family is supported by pedigree analysis of a further nine CEPH families, albeit with less dramatic examples, biased in part by the nature of the parental genotypes. Hence, we chose individuals with easily distinguishable allele sizes for analysis of sperm DNA. Analysis of matched sperm and blood DNA at BAT-40 by SP-PCR demonstrated a statistically significant increase in the proportion of mutant alleles in sperm compared to somatic DNA. This argues strongly that the mutations detected by SP-PCR of sperm DNA are indeed authentic. We employed rigorous controls to ensure against contamination (see Materials and Methods). The SP-PCR approach has been previously shown to detect mutant alleles with equal fidelity in sperm and constitutional DNA templates as demonstrated by direct comparisons between mutation rates detected by SP-PCR of sperm compared to family studies (11). These studies have consistently validated the SP-PCR approach (11,14).

The presence of an inactivating *MLH1* mutation in the germline of one individual did not further influence the level of instability at BAT-40 in the sperm DNA. Intriguingly, there were shorter mutant alleles predominating in the sperm DNA despite the fact that the SP-PCR technique reliably detected both large and short constitutional alleles.

The results presented here provide compelling evidence that BAT-40 is inherently highly unstable in the germline. It will be of interest to determine whether this phenomenon is common and what length of poly(A/T) tract represents a threshold at which instability becomes likely.

Hypermutability at the BAT-40 locus provides an explanation for the wide spectrum of allelic variants present in the

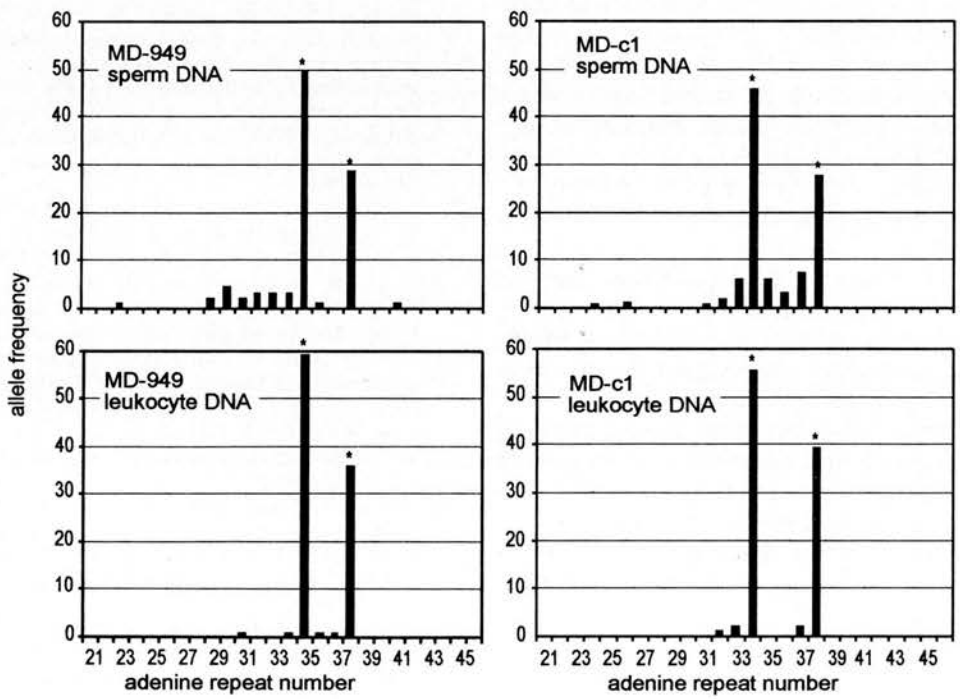


Figure 7. BAT-40 allele sizes in matched constitutional and sperm DNA detected by SP-PCR. The predominant allele sizes for each individual are indicated by asterisks, as detected from analysis of undiluted DNA. MD-949 is a CRC patient with a germline mutation in the human *MLH1* gene. MD-c1 is a normal healthy control individual.

Table 3. Summary of mutant alleles detected by SP-PCR in matched sperm and blood DNA from samples MD-949 and MD-c1

Sample	Total no. alleles	Mutants (frequency)
MD-949 Sperm	91	20 (0.22)
MD-949 Blood	99	4 (0.04) ($\chi^2 = 13.82$, $P < 0.001$)
MD-c1 Sperm	164	44 (0.27)
MD-c1 Blood	99	5 (0.05) ($\chi^2 = 19.32$, $P < 0.001$)

Scottish, CEPH and other populations studied, since transmission of new germline variants can become established within the population.

The evidence that BAT-40 represents a poly(A/T) tract within the genomic structure of a gene and exhibits instability in the germline might be of importance in understanding the mechanisms generating mutations at other such polymorphic repeat loci. Indeed this phenomenon may be common to many poly(A/T) tracts and further study of such sequences is merited to elucidate whether this is a widespread phenomenon. Poly(A/T) tracts are ubiquitous at the 3′UTR of all coding genes where the stability in length of the poly(A) tail is of known functional importance to the stability of the mRNA species (30). Of further relevance, these repetitive tracts are common in intronic sequence (17), and shortening of intronic mononucleotides has also been shown to have functional consequences. For instance, aberrant splice variants of the ATM gene that result in ataxia-telangiectasia can arise as a consequence of shortened intronic mononucleotide tracts (31). In addition the shortened $poly(T)_5$ variant in intron 8 of the cystic fibrosis transmembrane conductance regulator gene causes congenital bilateral absence of the vas deferens when associated with a cystic fibrosis mutation on the other allele (34). Mutation of poly(A/T) tracts within exonic sequences have also been shown to contribute to carcinogenesis and this is exemplified by mutation of the TGFBR2 gene in MSI^+ CRCs (20,21). Hence, it seems reasonable to speculate that the mechanism of inherent instability elucidated here might also have relevance to a number of genes containing such repeats.

BAT-40 is used routinely as a marker in determining tumour genomic stability in relation to defective DNA MMR, due to its extreme sensitivity to mutation in the absence of MMR activity (5,21,23). Microsatellite markers that display such germline hypermutability should be used with caution in view of the likelihood of mitotic instability. Very unstable markers may be too sensitive to provide the specificity to MMR defects that is clearly required in such screening strategies. The evidence reported here supports a growing number of studies that highlight the importance of understanding inherent characteristics influencing marker stability when they are used in clinical analyses (7,23,26,35).

ACKNOWLEDGEMENTS

This work was supported by the following grants: Cancer Research Campaign (SP2326/0201), Scottish Health Department (K/MRS/50/C2723) and Urquhart Charitable Trust. A.L.B. was funded by an MRC PhD Student Fellowship and S.M.F. by an RSE Personal Research Fellowship.

REFERENCES

1. Weissenbach,J., Gyapay,G., Dib,C., Vignal,A., Morissette,J., Millasseau,P., Vaysseix,G. and Lathrop,M. (1992) A second-generation linkage map of the human. *Nature*, **359**, 794–801.
2. Aaltonen,L.A., Peltomaki,P., Leach,F.S., Sistonen,P., Pylkkanen,L., Mecklin,J.P., Jarvinen,H., Powell,S.M., Jen,J. and Hamilton,S.R. (1993) Clues to the pathogenesis of familial colorectal cancer. *Science*, **260**, 812–816.
3. Thibodeau,S.N., Bren,G. and Schaid,D. (1993) Microsatellite instability in cancer of the proximal colon. *Science*, **260**, 816–819.
4. Ionov,Y., Peinado,M.A., Malkhosyan,S., Shibata,D. and Perucho,M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
5. Rodriguez-Bigas,M.A., Boland,C.R., Hamilton,S.R., Henson,D.E., Jass,J.R., Khan,P.M., Lynch,H., Perucho,M., Smyrk,T., Sobin,L. and Srivastava,S. (1997) A National Cancer Institute workshop on hereditary nonpolyposis colorectal cancer syndrome: meeting highlights and Bethesda guidelines. *J. Natl Cancer Inst.*, **89**, 1758–1762.
6. Dietmaier,W., Wallinger,S., Bocker,T., Kullmann,F., Fishel,R. and Ruschoff,J. (1997) Diagnostic microsatellite instability: definition and correlation with mismatch repair protein expression. *Cancer Res.*, **57**, 4749–4756.
7. Bacon,A.L., Farrington,S.M. and Dunlop,M.G. (2000) Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells. *Hum. Mol. Genet.*, **9**, 2707–2713.
8. Sturzeneker,R., Bevilacqua,R.A., Haddad,L.A., Simpson,A.J. and Pena,S.D. (2000) Microsatellite instability in tumors as a model to study the process of microsatellite. *Hum. Mol. Genet.*, **9**, 347–352.
9. Brinkmann,B., Klintschar,M., Neuhuber,F., Huhne,J. and Rolf,B. (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.*, **62**, 1408–1415.
10. Weber,J.L. and Wong,C. (1993) Mutation of human short tandem repeats. *Hum. Mol. Genet.*, **2**, 1123–1128.
11. Jeffreys,A.J., Tamaki,K., MacLeod,A., Monckton,D.G., Neil,D.L. and Armour,J.A. (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genet.*, **6**, 136–145.
12. May,C.A., Jeffreys,A.J. and Armour,J.A. (1996) Mutation rate heterogeneity and the generation of allele diversity at the human minisatellite MS205 (D16S309). *Hum. Mol. Genet.*, **5**, 1823–1833.
13. Monckton,D.G., Cayuela,M.L., Gould,F.K., Brock,G.J., Silva,R. and Ashizawa,T. (1999) Very large (CAG)(n) DNA repeat expansions in the sperm of two spinocerebellar ataxia type 7 males. *Hum. Mol. Genet.*, **8**, 2473–2478.
14. Crawford,D.C., Wilson,B. and Sherman,S.L. (2000) Factors involved in the initial mutation of the fragile X CGG repeat as determined by sperm small pool PCR. *Hum. Mol. Genet.*, **9**, 2909–2918.
15. Kunst,C.B., Leeflang,E.P., Iber,J.C., Arnheim,N. and Warren,S.T. (1997) The effect of FMR1 CGG repeat interruptions on mutation frequency as measured by sperm typing. *J. Med. Genet.*, **34**, 627–631.
16. Mornet,E., Chateau,C., Hirst,M.C., Thepot,F., Taillandier,A., Cibois,O. and Serre,J.L. (1996) Analysis of germline variation at the FMR1 CGG repeat shows variation in the normal-premutated borderline range. *Hum. Mol. Genet.*, **5**, 821–825.
17. Toth,G., Gaspari,Z. and Jurka,J. (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.*, **10**, 967–981.
18. International human genome sequencing consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
19. Markowitz,S. (2000) TGF-beta receptors and DNA repair genes, coupled targets in a pathway of human colon carcinogenesis. *Biochim. Biophys. Acta*, **1470**, M13–M20.
20. Markowitz,S., Wang,J., Myeroff,L., Parsons,R., Sun,L., Lutterbaugh,J., Fan,R.S., Zborowska,E., Kinzler,K.W. and Vogelstein,B. (1995) Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science*, **268**, 1336–1338.
21. Parsons,R., Myeroff,L.L., Liu,B., Willson,J.K., Markowitz,S.D., Kinzler,K.W. and Vogelstein,B. (1995) Microsatellite instability and mutations of the transforming growth factor beta type II receptor gene in colorectal cancer. *Cancer Res.*, **55**, 5548–5550.

22. Lachance,Y., Luu-The,V., Labrie,C., Simard,J., Dumont,M., de Launoit,Y., Guerin,S., Leblanc,G. and Labrie,F. (1990) Characterization of human 3 beta-hydroxysteroid dehydrogenase/delta 5-delta 4-isomerase gene and its expression in mammalian cells [published erratum appears in *J. Biol. Chem.* (1992), **267**, 3551]. *J. Biol. Chem.*, **265**, 20469–20475.
23. Boland,C.R., Thibodeau,S.N., Hamilton,S.R., Sidransky,D., Eshleman,J.R., Burt,R.W., Meltzer,S.J., Rodriguez-Bigas,M.A., Fodde,R., Ranzani,G.N. and Srivastava,S. (1998) A National Cancer Institute workshop on microsatellite instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. *Cancer Res.*, **58**, 5248–5257.
24. Liu,B., Farrington,S.M., Petersen,G.M., Hamilton,S.R., Parsons,R., Papadopoulos,N., Fujiwara,T., Jen,J., Kinzler,K.W., Wyllie,A.H., Vogelstein,B. and Dunlop,M. (1995) Genetic instability occurs in the majority of young patients with colorectal cancer. *Nature Med.*, **1**, 348–352.
25. Zhou,X.P., Hoang,J.M., Cottu,P., Thomas,G. and Hamelin,R. (1997) Allelic profiles of mononucleotide repeat microsatellites in control individuals and in colorectal tumours with and without replication errors. *Oncogene*, **15**, 1713–1718.
26. Samowitz,W.S., Slattery,M.L., Potter,J.D. and Leppert,M.F. (1999) BAT-26 and BAT-40 instability in colorectal adenomas and carcinomas and germline polymorphisms. *Am. J. Pathol.*, **154**, 1637–1641.
27. Jeffreys,A.J., Neumann,R. and Wilson,V. (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. *Cell*, **60**, 473–485.
28. Vilkki,S., Tsao,J.L., Loukola,A., Poyhonen,M., Vierimaa,O., Herva,R., Aaltonen,L.A. and Shibata,D. (2001) Extensive somatic microsatellite mutations in normal human tissue. *Cancer Res.*, **61**, 4541–4544.
29. Yokozaki,H. (2000) Distribution of germline BAT-40 poly-adenine tract microsatellite variants in the Japanese. *Int. J. Mol. Med.*, **6**, 445–448.
30. Bernstein,P. and Ross,J. (1989) Poly(A), poly(A) binding protein and the regulation of mRNA stability. *Trends Biochem. Sci.*, **14**, 373–377.
31. Ejima,Y., Yang,L. and Sasaki,M.S. (2000) Aberrant splicing of the ATM gene associated with shortening of the intronic mononucleotide tract in human colon tumour cell lines: a novel mutation target of microsatellite instability. *Int. J. Cancer*, **86**, 262–268.
32. Chu,C.S., Trapnell,B.C., Curristin,S., Cutting,G.R. and Crystal,R.G. (1993) Genetic basis of variable exon 9 skipping in cystic fibrosis transmembrane conductance regulator mRNA. *Nature Genet.*, **3**, 151–156.
33. Bacon,A.L., Farrington,S.M. and Dunlop,M.G. (2001) Mutation frequency in coding and non-coding repeat sequences in mismatch repair deficient cells derived from normal human tissue. *Oncogene*, **20**, in press.
34. Chillon,M., Casals,T., Mercier,B., Bassas,L., Lissens,W., Silber,S., Romey,M.C., Ruiz-Romero,J., Verlingue,C. and Claustres,M. (1995) Mutations in the cystic fibrosis gene in patients with congenital absence of the vas deferens. *N. Engl. J. Med.*, **332**, 1475–1480.
35. Frazier,M.L., Sinicrope,F.A., Amos,C.I., Cleary,K.R., Lynch,P.M., Levin,B. and Luthra,R. (1999) Loci for efficient detection of microsatellite instability in hereditary non-polyposis colorectal cancer. *Oncol. Rep.*, **6**, 497–505.

Oncogene (2001) 20, 7464–7471

www.nature.com/onc

Mutation frequency in coding and non-coding repeat sequences in mismatch repair deficient cells derived from normal human tissue

Andrea L Bacon[1], Susan M Farrington[1] and Malcolm G Dunlop*,[1]

[1]*University of Edinburgh Department of Oncology and MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK*

Repetitive tracts within the coding regions of *TGFBR2* and *BAX* are frequently mutated in mismatch repair deficient tumours and are implicated in tumour progression. However, there has been little study of the balance between selection pressure and inherent instability at sequences within these genes. To determine whether *TGFBR2* and *BAX* are inherently prone to mutations in the presence of MMR defects, we studied MMR deficient cells derived from B-lymphocytes. By analysing cells derived from normal tissue we aimed to minimize the effects of selection pressures that bias the apparent frequency of mutation. We definitively show that certain sequences, usually repaired by MMR, are inherently unstable. Using a small pool PCR technique we confirmed these cells exhibit microsatellite instability. Additionally, we demonstrate that MMR deficiency results in an excess of mutations, specifically at the poly(A)$_{10}$ tract compared to other regions of the *TGFBR2* gene ($P<0.001$). Conversely, an excess of mutations does not appear to arise at the poly(G)$_8$ tract of the *BAX* gene. These studies provide insight into the mechanism by which *TGFBR2* and *BAX* genes become mutated during tumorigenesis. These findings invoke the notion of 'unmasking' specific hypermutable sequences in particular genes adding further complexity to the concept of the mutator phenotype. *Oncogene* (2001) **20,** 7464–7471.

Keywords: TGFBR2; mutation rate; BAX; mismatch repair; colon cancer

Introduction

A mutator phenotype is well documented in tumours of patients with hereditary non-polyposis colorectal cancer (HNPCC), and also in a subset of sporadic cancers (Aaltonen *et al.*, 1993). Such genomic instability is largely due to inactivation of one or more of the mismatch repair (MMR) genes (*MLH1*, *MSH2*, *PMS2*, *PMS1*, *MSH6*) (Papadopoulos and Lindblom, 1997) by mutation or promoter hypermethylation (Wheeler *et al.*, 1999). Disruption of MMR leads to accumulation of mutations at simple microsatellite sequences and within the repetitive tracts of coding sequence (Ionov *et al.*, 1993; Parsons *et al.*, 1995a). A mutator phenotype, combined with selection pressure, is believed to allow the accumulation of mutations in key genes leading to progression of neoplasia from cellular clones to invasive cancers (Cahill *et al.*, 1999). However, the relative contribution of selection pressure and inherent instability at specific gene sequences is not clear.

Mutations in repetitive sequences within a number of genes have been identified in microsatellite unstable (MSI$^+$) tumours, including *TGFBR2*, *BAX* and *IGFR2*, (Markowitz *et al.*, 1995; Ouyang *et al.*, 1997; Rampino *et al.*, 1997). However, the observation of gene mutations in MMR deficient cancers does not prove that such mutations are tumorigenic and not simply bystander effects. Other lines of evidence are required to establish pathogenicity (Boland *et al.*, 1998).

Mutations in the poly(A)$_{10}$ tract of the transforming growth factor beta type 2 receptor (*TGFBR2*) gene are observed in over 90% of MMR deficient colorectal tumours (Markowitz *et al.*, 1995). The transforming growth factor beta (TGFB) signalling pathway results in potent anti-proliferative responses (Wang *et al.*, 1995). Loss of this negative regulatory mechanism results in excess cell growth suggesting that the *TGFBR2* gene behaves as a tumour suppressor. Bi-allelic inactivation of *TGFBR2* has been demonstrated in most MSI$^+$ colorectal cancers (CRC's) (Parsons *et al.*, 1995a). Furthermore, in microsatellite stable (MSS) CRC's, inactivating point mutations in *TGFBR2* (Grady *et al.*, 1999) and, mutational inactivation of other components of the TGFB signalling pathway (SMAD2 and SMAD4), have also been observed (Eppert *et al.*, 1996). Together such studies suggest inactivation of TGFBR2 and subsequent disruption of the TGFB signalling pathway is an important step during colon carcinogenesis.

Around 50% of MSI$^+$ cancers contain mutations in the poly(G)$_8$ tract of the pro-apoptotic *BAX* gene (Rampino *et al.*, 1997). Bi-allelic inactivation in some cases suggests that *BAX* mutations are important during progression of MMR deficient tumours. How-

*Correspondence: MG Dunlop, Colon Cancer Genetics Group, MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, EH4 2XU, UK;
E-mail: Malcolm.Dunlop@hgu.mrc.ac.uk
Received 21 June 2001; revised 5 September 2001; accepted 13 September 2001

ever, inactivation of both alleles is not observed with the same frequency as in the *TGFBR2* gene (Rampino *et al.*, 1997). The prevalence of mutations observed in the repetitive tracts of both *TGFBR2* and *BAX* are in excess of those at non-coding repeats in intronic sequence, suggesting a clonal selection effect (Zhang *et al.*, 2001).

Mutations at microsatellites occur consistently in adenomas with MMR gene defects (Shibata *et al.*, 1994) and have also been detected in the earliest lesions known as aberrant crypt foci (Pedroni *et al.*, 2001). Such observations suggest that microsatellite instability (MSI) is an early event in the progression of a normal cell towards a cancerous phenotype. *TGFBR2* mutations are similarly observed within MMR deficient adenomas suggesting that these mutations also occur early during tumorigenesis (Grady *et al.*, 1998; Abdel-Rahman *et al.*, 1999). Although *BAX* mutations have been observed in early tumours, there are also data suggesting they occur at a later stage in tumour development (Abdel-Rahman *et al.*, 1999).

The relative contribution of inherent, sequence-specific instability has not been studied extensively. Early neoplastic lesions are subject to selection pressure and so evolve clonally, analogous to more advanced tumours. Furthermore, even early lesions have accumulated multiple mutations making it difficult to dissect the contribution of events resulting directly and exclusively from MMR defects.

We hypothesized that differences in mutational frequencies at the *TGFBR2* and *BAX* genes may reflect differences in the relative contributions of inherent mutation and effects of selection at these loci. Therefore we set out to investigate the intrinsic instability in these genes using a system where MMR is defective but cells are not subject to bias due to molecular changes that characterize the malignant phenotype. We conducted a series of studies in lymphoblastoid cells that are deficient in MMR, and we show sequence specific hypermutability at the poly(A)$_{10}$ tract of the *TGFBR2* gene arising in cells of normal lineage.

Results

Characterization of the mutator phenotype in cell lines derived from lymphocytes with constitutive mismatch repair defects

Lymphoblast cell lines lbl-1261 and lbl-1260 are derived from lymphocytes of patients with germline mutations of the MMR genes *PMS2* and *MLH1* respectively (Parsons *et al.*, 1995b). These mutations have both previously been demonstrated to result in the complete loss of MMR activity (Parsons *et al.*, 1995b). The *PMS2* mutation in lbl-1261 exerts a dominant negative effect (Nicolaides *et al.*, 1998), while we have shown that MLH1 expression is dramatically reduced in lbl-1260 (data not shown).

We set out to characterize in detail the mutator phenotype previously reported in these cell lines (Parsons *et al.*, 1995b). Both cell lines were confirmed to be derived from B-lymphocytes by demonstration of the appropriate pattern of expression of lymphocyte and epithelial cell specific markers, detected by flow cytometry (data not shown).

We anticipated that mutations in these cell lines were likely to arise at a low frequency within a population of predominantly wild type cells. Hence, we employed a sensitive, small pool PCR (SP-PCR) strategy to genotype individual microsatellite alleles from lbl-1261 and lbl-1260 DNA plus control cell lines (lbl-c5 and lbl-c1) at two microsatellite markers (BAT-40 and D2S123) (Figure 1). Cell lines lbl-1260 and lbl-1261 both displayed a substantially greater proportion of mutant alleles compared with control cell lines (Figure 1 and Table 1) in accordance with previous findings (Parsons *et al.*, 1995b). For lbl-1261, 89/270 (33%) D2S123 alleles and 71/139 (51.1%) BAT-40 alleles were mutant, as were 10/120 (8.3%) and 13/54 (24.1%) alleles typed from lbl-1260. The proportion of mutant alleles for lbl-1261 and lbl-1260 was significantly greater than that observed in control cell lines, at both BAT-40 ($\chi^2=50.5$, $P<0.001$; $\chi^2=15.4$; $P<0.001$) and D2S123 ($\chi^2=28.7$, $P<0.001$; $\chi^2=4.4$, $P=0.036$) (Table 1). Notably lbl-1261 displayed a significantly higher level of microsatellite instability compared to lbl-1260 ($\chi^2=25.2$, $P<0.001$). Insertion and deletion mutant alleles were observed indicating that PCR bias for reductions in tract length was not a major factor.

There was evidence of heterogeneity in the degree of instability between microsatellite loci. The proportion of mutant BAT-40 alleles was significantly greater than that of D2S123 in both cell lines ($\chi^2=9.9$, $P<0.001$) suggesting differential susceptibility of simple sequence repeats (SSRs) to replication error.

These data show firstly, that lymphoblastoid cell lines lbl-1260 and lbl-1261 with *MLH1* and *PMS2* mutations respectively, exhibit microsatellite instability. Secondly, that there is considerable heterogeneity in the degree of instability between the two MMR deficient cell lines, presumably due to the MMR gene mutation itself. Furthermore, the results demonstrate inherent differences in instability between particular SSR's.

Excess mutations arise in the poly(A)$_{10}$ tract of TGFBR2 *in lbl-1261*

To address whether the mutator phenotype described above also results in instability within genes known to be involved in colorectal carcinogenesis, we analysed lbl-1261 and lbl-1260 for mutations in the poly(A)$_{10}$ tract of *TGFBR2*. There were considerable difficulties in reproducibly detecting mutations in alleles of the *TGFBR2* poly(A)$_{10}$ tract amplified by SP-PCR using LoVo as a positive control. Therefore we adapted a sensitive assay to detect 1 bp deletions occurring at low frequency within the poly(A)$_{10}$ tract of *TGFBR2* (Mironov *et al.*, 1999). In brief, the reverse primer introduces a *Hinf* I site in the presence of a 1 bp deletion in the poly(A)$_{10}$ tract, whereas in the presence of wild type sequence no *Hinf* I restriction site is

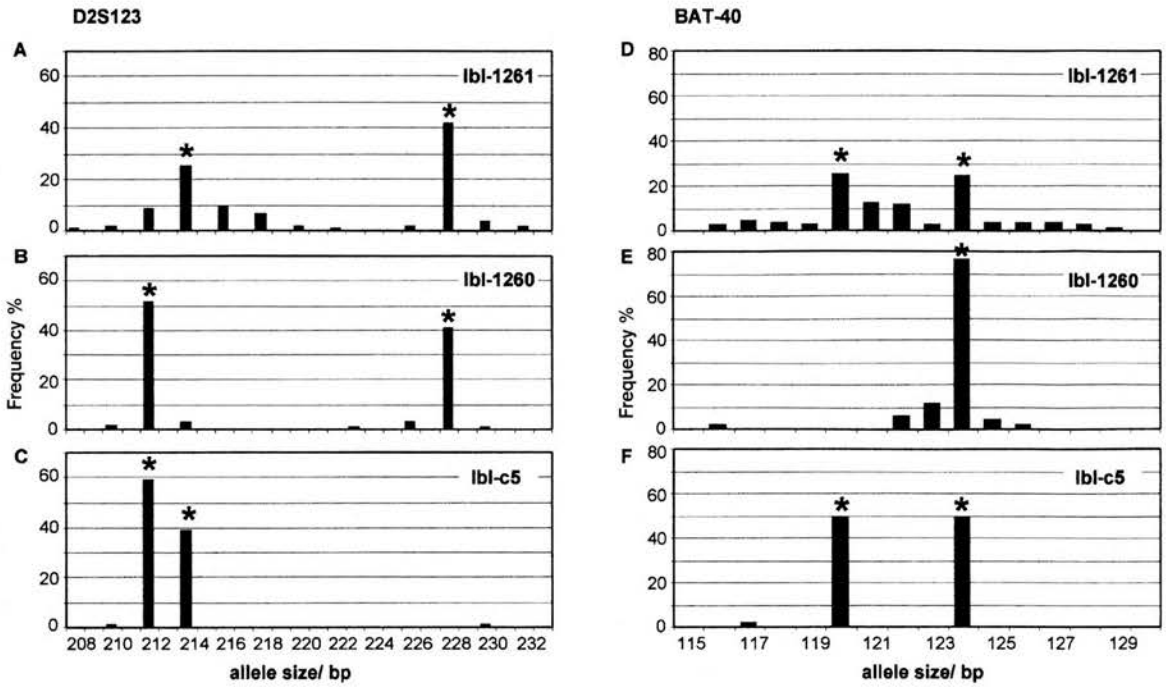


Figure 1 D2S123 (**a**–**c**) and BAT-40 (**d**–**f**) alleles identified in MMR deficient (lbl-1261 (**a** and **d**) and lbl-1260 (**b** and **e**)) and MMR proficient (lbl-c5 (**c** and **f**)) lymphoblastoid cell lines detected by SP–PCR. The predominant constitutional alleles (*) were sized by analysis of undiluted DNA. Alleles for D2S123 (**a** and **c**) have been reported by us previously (Bacon *et al*., 2000) but are shown here for comparison

Table 1 Summary data of alleles typed at DS2123 and BAT-40 SSR loci in lbl-1261, lbl-1260 and controls. BAT-40 control allele data is composed of pooled allele counts for lbl-c5 and second cell line lbl-c1 that was also typed at BAT-40

	DS2123		*BAT-40*	
Cell line	*Total alleles*	*Mutants* (% alleles typed)	*Total alleles*	*Mutants* (% alleles typed)
lbl-1261	270	89 (33.0)	139	71 (51.1)
lbl-1260	120	10 (8.3)	54	13 (24.1)
Controls	109	2 (1.8)	154	6 (3.9)

introduced during the PCR step (Mironov *et al*., 1999). Digestion products are detected using an ABI310 genetic analyser. This assay reliably detected the presence of mutant *TGFBR2* alleles in positive control lines, LoVo and HCT116 (Figure 2). Mutations were also detected in lbl-1261 DNA, indicating an appreciable level of mutant alleles in lbl-1261 (Figure 2). Intriguingly, the frequency of mutant alleles indicated for lbl-1260 was not significantly different to that of wild type control DNAs (Figure 2).

In order to quantify and to characterize mutations arising in lbl-1261, exon 3 of *TGFBR2* was PCR cloned (Figure 3 and Table 2). Of 56 clones sequenced, 17 (30.3%) were mutated at the poly(A)$_{10}$ tract. In contrast only three (6.8%) of the 44 wild type control clones were mutant presumably representing technical artefact. These data demonstrate a significant excess of mutations at the poly(A)$_{10}$ region of *TGFBR2* in lbl-1261 compared to a control cell line ($\chi^2=8.5$, $P=0.003$). Most of the exon 3 poly(A)$_{10}$ mutations detected in lbl-1261 were 1 bp deletions (82%), supporting the validity of the results from the digest assay described above. However a small number of 1 bp insertions, plus an A→G transition were also identified. The frequency and the spectrum of these mutations strongly suggests they are not constitutional variants but arise somatically. These data provide compelling evidence that *de novo* mutations arise frequently due to inherent instability of the poly(A)$_{10}$ repeat in *TGFBR2* but these are normally repaired by a proficient DNA MMR apparatus.

Determination of the relative susceptibility to mutation of regions within TGFBR2

Next we wished to determine the relative susceptibility to mutation of regions within the *TGFBR2* gene. In particular to address whether the mutator phenotype results in predilection for mutation at the poly(A)$_{10}$ repeat sequence or whether the observations were simply the result of non-specific increase in mutation rate. Therefore we compared mutant allele counts arising at non-repeat regions of the *TGFBR2* gene in lbl-1261 to those arising in the poly(A)$_{10}$ tract. A 216 bp region of exon 3 flanking the repetitive poly(A)$_{10}$ region and a 245 bp non-repeat region of exon 4 were screened for mutations by PCR cloning and sequencing of individual alleles. Exon 4 contains a site found to be mutated in MSS colorectal cancer (Lu *et al*., 1998) but lacks long repeats. No mutations were identified in this region in a total of 7840 bp of sequence from 32 lbl-1261 alleles sequenced for exon 4

(Figure 4 and Table 2). However, in the non-repetitive region of exon 3, six transition mutations were identified in a total of 10928 bp of sequence analysed from 49 lbl-1261 alleles, a frequency of 0.55 mutations/kb (Figure 4 and Table 2). There were no mutations in the non-repetitive region of exon 3 in a total of 8897 bp from 40 control (lbl-c5) alleles (Figure 4). These data indicate that exon 3 of the *TGFBR2* gene appears particularly prone to mutation, compared with exon 4 of the same gene ($P=0.036$), and also compared to the same region in a control cell line ($P=0.036$). Nonetheless, the proportion of mutations detected in the non-repeat region of *TGFBR2* is significantly lower than that for the poly(A)$_{10}$ tract ($P<0.001$).

Taken together, these results show that exon 3 of *TGFBR2* is inherently mutable and that predilection for instability at the poly(A)$_{10}$ tract contributes to the frequent observation of *TGFBR2* mutations in MSI$^+$ tumours.

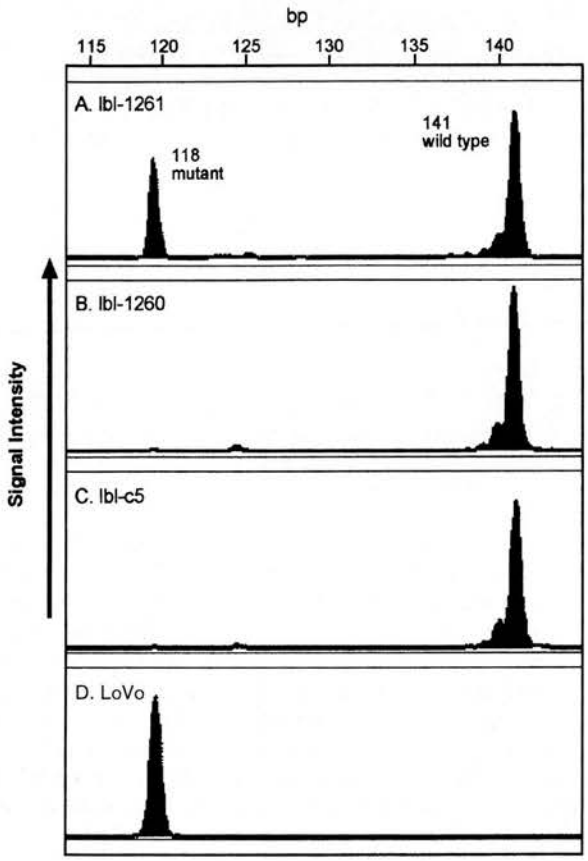


Figure 2 Representative ABI 310 traces of the poly(A)$_{10}$ tract in exon 3 of the *TGFBR2* gene from wild type (lbl-c5) (**c**) and mutant (LoVo) (**d**) control cell lines corresponding to the presence or absence of mutant template alleles after *Hinf* I digestion. There were only constitutional genotype peaks for lbl-1260 (**b**), but mutant peaks were reproducibly detected for lbl-1261 (**a**)

Analysis of inherent instability at the BAX *gene in MMR deficient cells*

Since lbl-1261 was particularly unstable at *TGFBR2* and non-coding repeats, we wished to determine whether this phenotype was associated with a high level of mutation in the poly(G)$_8$ repeat of the pro-apoptosis gene, *BAX*. SP–PCR was employed to genotype alleles using LoVo as a positive control. Of 164 SP–PCR products only two mutants (1.2%) were detected by DHPLC analysis of the 94 bp products (data not shown). Sequencing of mutant SP–PCR products confirmed a 1 bp insert at the poly(G)$_8$ tract. No mutations were detected in 106 SP–PCR products analysed from a control cell line. These data suggest that mutations do also occur at the repetitive poly(G)$_8$ tract of the *BAX* gene, but the frequency is much lower

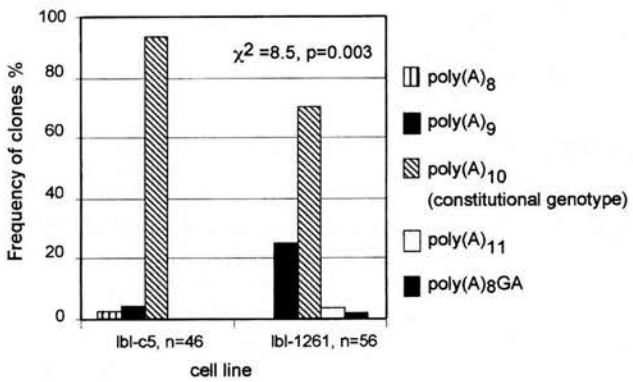


Figure 3 Summary data showing an excess of mutations in the poly(A)$_{10}$ tract of *TGFBR2* in lbl-1261 as detected by allele cloning and sequencing ($P=0.003$)

Table 2 Summary of mutations detected, as a function of total bps sequenced, at three separate regions within the *TGFBR2* gene in lbl-1261 and control cells

Region	TGFBR2 *exon 3 poly(A)$_{10}$ repeat*		TGFBR2 *exon 3 non-repeat*		TGFRB2 *exon 4*	
analysed	*(nt 709 to nt 718, 10 bp)*		*(nt 599-10 to nt 789, 216 bp)*		*(nt 1021 to nt 1266, 245 bp)*	
cell lines	*lbl-1261*	*Control*	*lbl-1261*	*Control*	*lbl-1261*	*Control*
Total sequence analysed/Kb	0.56	0.44	11.016	8.967	7.840	ND
Mutations identified	14×1 bp del 2×1 bp ins 1×a-g transition	2×1 bp del 1×2 bp del	t-c nt 744 a-g nt 602 a-g nt 677 a-g nt 702 t-c nt 725 c-t nt 651	None	None	ND
Mutation frequency, Mutations/Kb	30.36	6.82	0.55	Undetectable	Undetectable	ND

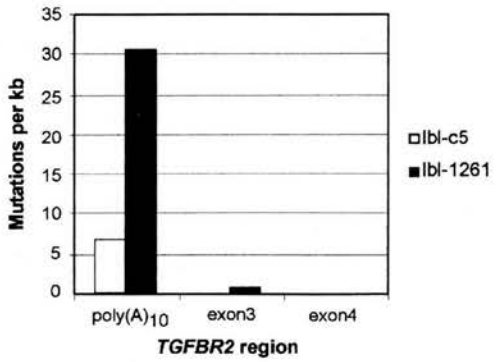


Figure 4 Comparison of the non-repeat coding sequences to the poly(A)$_{10}$ tract in *TGFBR2*. Mutation data are presented as mutations/kb sequenced, to account for variation in the length of cloned fragments

than that arising in *TGFBR2* and is below the level of reliable detection.

Discussion

The data presented here demonstrate a substantial level of inherent instability at SSRs and in coding sequences as a consequence of MMR deficiency in cell lines lbl-1260 and lbl-1261 that are derived from non-tumour tissue. By analysis of alleles by SP–PCR we have demonstrated that despite being of B-lymphocyte lineage, lbl-1260 and lbl-1261 exhibit an appreciable level of microsatellite instability, confirming previous findings (Parsons *et al*., 1995b). We extend these findings and show that there is significant heterogeneity with regards to both the spectrum and frequency with which mutations occur at microsatellite sequences in these cells. We also provide evidence that inherent mutability in the *TGFBR2* gene contributes to the frequent observation of mutations in MSI$^+$ tumours. The instability shows a clear propensity to the same poly(A)$_{10}$ tract that is frequently mutated in MSI$^+$ tumours. Our analysis of the poly(G)$_8$ tract of the *BAX* gene indicates that there are varying degrees of inherent mutational instability within coding regions known to have a high frequency of mutation in tumours, and suggests differential contributions from mutational mechanisms and effects of the tumour phenotype itself.

We observed evidence of heterogeneity of instability at microsatellite markers D2S123 and BAT-40, in lbl-1260 and lbl-1261. Heterogeneity in mutation frequency between BAT-40 and D2S123 indicates locus specific influences, and suggests BAT-40 is an inherently more unstable locus (Bacon *et al*., 2001). Intrinsic sequence characteristics at given loci can influence the manifestation of the mutator phenotype and the mutation bias between alleles at D2S123 has been previously reported (Bacon *et al*., 2000).

We employed two independent experimental approaches using MMR deficient cells to demonstrate that the poly(A)$_{10}$ tract of *TGFBR2* is intrinsically hypermutable even in cells that are not tumour derived. The data show that mutations can be detected as a consequence of MMR deficiency when the effects of selection pressure and other confounding molecular variables present in tumour cells, are minimized. These studies were performed in cell lines that represent a model assay system and as such, selection cannot be entirely negated. EBV transformed lymphoblast cell lines are resistant to the effects of TGFB1 due to selection against TGFBR2 expression (Inman and Allday, 2000). However, it has been shown that this is not due to mutation of the poly(A)$_{10}$ repeat in exon 3 (Inman and Allday, 2000). Furthermore, the comparison with control lines shows conclusively there is an important MMR dependent effect.

Since we detected a number of different *TGFBR2* mutant alleles, the findings are consistent with the occurrence of somatic events rather than pre-existing constitutional heterozygous mutations or mosaicism. The fact that excess mutations were not observed in *TGFBR2* in lbl-1260 might indicate a threshold effect of MMR activity in the cell lines studied here, which has an influence on the ability for replication errors to be repaired.

The fact that we detected mutations in sequences encoding tumour suppressor genes as well as non-coding repeats in cells derived from a normal B-cell lineage is remarkable in itself. That lbl-1261 appears particularly prone to mutation may be due to gene specific influences. Indeed, a recent report (Vilkki *et al*., 2001) did not detect any *TGFBR2* mutant alleles in autopsy material from a child with homozygous *MLH1* mutations. Interestingly, cell line lbl-1260 which contains a heterozygous *MLH*1 mutation displays a lower incidence of mutation at both coding and non-coding sequence (Parsons *et al*., 1995b).

Few studies have addressed mutation frequency at the non-repetitive sequences surrounding the mutable repeat tracts in genes implicated in tumorigenesis in MSI$^+$ tumours (Takenoshita *et al*., 1997). Here we demonstrate that non-repeat regions are not subject to high levels of mutation, indicating that it is the repeat sequence and not the entire *TGFBR2* gene, which is prone to instability consequent of MMR deficiency.

Analysis of the poly(G)$_8$ repeat in the *BAX* gene, suggests that although this is mutated frequently in MMR deficient colorectal tumours, it does not share the same intrinsic propensity for mutation observed at the poly(A)$_{10}$ repeat of *TGFBR2*. These results give considerable insight into susceptibility of the *BAX* and *TGFBR2* genes to mutations arising exclusively as a result of MMR defects. The high mutability of the poly(A)$_{10}$ tract is likely to be a consequence of increased replication error at this region. Such error is well documented and, dependent on the number of repeats in the tract (Parsons *et al*., 1995a; Dietmaier *et al*., 1997). Chromatin structure within or surrounding such sequences has also been suggested to contribute to mutability at repeat tracts (Zhang *et al*., 2001). In this regard it is interesting to note that we detected a low, but appreciable, frequency of mutation in the flanking

sequence surrounding the poly(A)$_{10}$ tract in lbl-1261 that was significantly greater than that observed for exon 4 of the same gene.

That the *TGFBR2* poly(A)$_{10}$ tract is so readily prone to mutation goes some way to explain why these mutations are observed with such a consistently high frequency within MMR deficient cancers (Markowitz *et al*., 1995) and also in early adenomas (Abdel-Rahman *et al*., 1999). Taken together with these previous studies, the data presented here supports the notion that MMR inactivation occurs very early in neoplastic transformation and results in accumulation of mutations in the *TGFBR2* poly(A)$_{10}$ tract. Subsequently, selection pressure for homozygous mutations would be anticipated to result in clonal selection within the tumour. Since we detected only a very low level of *BAX* mutations in the most unstable line lbl-1261, this suggests that *BAX* is inherently more stable and that selection pressures may play a greater role than is the case for *TGFBR2* mutations. The data presented here suggest that *BAX* mutations arise infrequently and this may explain why they are less frequently observed or are not detectable at such an early stage in tumorigenesis as *TGFBR2* mutations (Rampino *et al*., 1997; Abdel-Rahman *et al*., 1999).

Taken together these findings indicate that intrinsic mutational instability is an important determinant of mutation frequency observed in MSI tumours. Use of cells derived from normal tissue has allowed analysis of the mutation frequency consequent upon MMR deficiency, while minimizing bias from the effects of selection and confounding abnormalities. It will be of interest to determine which other coding gene sequences are similarly susceptible to mutation in cell lines lbl-1260 and lbl-1261 as this might identify important genes involved in MMR dependent tumour initiation and progression.

Materials and methods

Lymphoblast cell lines

Epstein-Barr virus (EBV) transformed lymphocytes from two healthy control individuals (lbl-c5, lbl-c1) were cultured in RPMI with 10% FCS 1% P/S. EBV transformed lymphocytes from two patients with colorectal cancer and phenotypic evidence of Turcots syndrome (lbl-1261, lbl-1260) were a gift from Bert Vogelstein. These patients carry germline defects in *PMS2* and *MLH1* respectively and have been shown to lack MMR activity (Parsons *et al*., 1995b).

Flow cytometry

Cells were analysed for the presence of B-cell specific and epithelial cell specific markers. Antibodies used were PE conjugated anti-human CD19 (B-cell specific) (Caltag) and, unconjugated anti-cytokeratin antibody (epithelial cell specific) (Clone BER-EP4, DAKO) detected using a second step FITC conjugated anti-mouse IgG. Data for 10 000 cells was acquired and analysed using a FACSCaliber and Cell Quest Software (Becton Dickinson). Controls included were unstained cells and cells stained with FITC conjugated anti-mouse IgG alone. The cancer cell line HCT116 provided a positive control for the epithelial marker BER-EP4.

Small pool PCR

DNA from cell lines was diluted to a final concentration of 15–20 pg per PCR reaction as previously described (Bacon *et al*., 2000). Fluorescently labelled D2S123 and BAT-40 primers (Parsons *et al*., 1995a) were used for microsatellite analysis. Non fluorescent primers were used to amplify a 94 bp region encompassing the poly(G)$_8$ repeat at codons 38 to 41 of the BAX gene (Rampino *et al*., 1997).

High fidelity PCR amplifications were performed using Expand high fidelity PCR system as previously described (Bacon *et al*., 2000). Reactions were prepared in 96 well plates. DNA free controls were prepared in 16 wells per plate and positive controls containing 100 ng of cell line DNA were prepared in two wells in every plate. Amplification was performed on an Omnigene PCR system thermal cycler (Hybaid) at 94°C (3 min) for 1 cycle, 94°C (1 min), 55°C (1 min), 72°C (1 min) for 35 cycles, 72°C (5 min) for 1 cycle. For microsatellite markers 2 μl of each PCR reaction including positive and negative controls was analysed on an ABI310 Automated Genetic Analyser, using Genescan software. For the *BAX* gene ABI310 analysis was not sensitive enough to pick up a 1 bp deletion in a known mutant, LoVo (Carethers and Pham, 2000), due to the small size of the fragment. 3 μl of BAX SP–PCR products were subjected to a second round of PCR allowing visualization of products on an agarose gel. These were subsequently analysed by DHPLC using a Transgenomic wave^{TM} machine along with the known mutant control.

The frequency of mutant alleles in each cell line for both microsatellite markers was expressed as the number of alleles which were mutant in length, divided by the total number of alleles detected (normal and mutant). Accordingly, frequencies are not exact contents of cells with alterations but relative values. Differences between mutation frequency in the two cell lines compared to controls, and between MSI frequency at the two microsatellite markers was evaluated using a Chi-squared test (Minitab V.13). Significance was taken at the 5% level.

TGFBR2 *restriction digest mutation detection assay*

We adapted a sensitive assay (Mironov *et al*., 1999) to provide an initial analysis to detect 1 bp deletions in the poly(A)$_{10}$ tract of *TGFBR2*. DNA samples from lbl-1261, lbl-1260, control wild type (lbl-c5) and a control mutant cell lines, LoVo and HCT116 (Carethers and Pham, 2000) were amplified using a primer that induces a *Hinf* I restriction site in the presence of this mutation. The forward primer was fluorescently labelled.

Forward primer: CACTCTAGGAGAAAGAATGACG.
Reverse primer: GAAAGTCTCACCAGGCTTTTTGATT.

These primers would be expected to amplify alleles of length A_9 and longer. Since LoVo contains one A_9 and one A_8 allele (Carethers and Pham, 2000) the primers used would not be expected to amplify the A_8 allele. PCR reactions were performed in a final volume of 25 μl using the Expand high fidelity PCR system as previously described (Bacon *et al*., 2000 and above). Five μl of PCR product was digested at 37°C overnight in a total volume of 15 μl with 5 units *Hinf*I restriction enzyme and 1X buffer H (Boehringer Mannheim).

Products were analysed on an ABI 310 Automated Genetic Analyser, using Genescan software. Wild type undigested fragments are visualized at 141 bp and mutant *Hinf* I digest fragments at 118 bp. The analysis was performed on each cell line in triplicate.

Cloning and sequence analysis

A 266 bp region of *TGFBR2* exon3 encompassing the poly$(A)_{10}$ tract was PCR cloned from lbl-1261 and wild type control, lbl-c5, DNA (primers described elsewhere, Lu *et al*., 1996). A 245 bp non-repetitive region of *TGFBR2* exon 4 was also cloned from lbl-1261 DNA. PCR reactions were performed as above using the following primers;

Forward: CCACGTGTGCCAACAACATCAACC
Reverse: CAGCCGTCAGGAACTGGAGTA

PCR fragments were cloned into TA cloning vectors using TOPO TA Cloning kit version K2 (Invitrogen BV, Groningen, The Netherlands) according to the manufacturer's instructions. Plasmid DNA was extracted from transformants using a QIAprep miniprep kit (QIAGEN Ltd, Crawley, UK). Sequencing was performed using M13 primers and PRISM ready Dye Terminator Cycle Sequencing Kit with AmpliTaq DNA polymerase, FS (Taq-FS: Perkin Elmer/Applied Biosystems) on an Applied Biosystems DNA sequencer model 373A or 377. DNA sequence analysis was performed by use of Sequencing analysis 3.0 and Sequencher^{TM} 3.0 (Gene Codes Corporation). All clones were sequenced at least in duplicate and the majority of mutants were re-confirmed by re-isolating and sequencing clone DNA from bacterial stocks.

A two tailed fishers exact test (http://home.clara.net/sisa/fisher.htm) with significance taken at the 5% level was used to evaluate differences in the number of mutations observed in the total amount of cloned sequences from different loci, or the same loci between different cell lines. When using this method it is assumed that mutations occur independently at any given base pair within the given locus.

Acknowledgments

We are grateful to Bert Vogelstein for the gift of cell lines lbl-1260 and lbl-1261. We acknowledge technical assistance from Sheila McBeath and assistance with flow cytometry from Martin Waterfall. This work was supported by the following grants; Cancer Research Campaign (SP2326/0201), Scottish Health Department (K-/MRS/50/C2723), The Melville Trust, Edinburgh University Cancer Research Endowment Fund and Urquart Charitable Trust. AL Bacon was funded by MRC PhD Student Fellowship, SM Farington by an RSE Personal Research Fellowship.

References

Aaltonen LA, Peltomaki P, Leach FS, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Powell SM, Jen J and Hamilton SR. (1993). *Science*, **260,** 812–816.

Abdel-Rahman WM, Georgiades IB, Curtis LJ, Arends MJ and Wyllie AH. (1999). *Oncogene*, **18,** 2139–2142.

Bacon AL, Dunlop MG and Farrington SM. (2001). *Nucleic Acids Res.*, in press.

Bacon AL, Farrington SM and Dunlop MG. (2000). *Hum. Mol. Genet.*, **9,** 2707–2713.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN and Srivastava S. (1998). *Cancer Res.*, **58,** 5248–5257.

Cahill DP, Kinzler KW, Vogelstein B and Lengauer C. (1999). *Trends Cell Biol.*, **9,** M57–M60

Carethers JM and Pham TT. (2000). *In Vivo*, **14,** 13–20.

Dietmaier W, Wallinger S, Bocker T, Kullmann F, Fishel R and Ruschoff J. (1997). *Cancer Res.*, **57,** 4749–4756.

Eppert K, Scherer SW, Ozcelik H, Pirone R, Hoodless P, Kim H, Tsui LC, Bapat B, Gallinger S, Andrulis IL, Thomsen GH, Wrana JL and Attisano L. (1996). *Cell*, **86,** 543–552.

Grady WM, Rajput A, Myeroff L, Liu DF, Kwon K, Willis J and Markowitz S. (1998). *Cancer Res.*, **58,** 3101–3104.

Grady WM, Myeroff LL, Swinler SE, Rajput A, Thiagalingam S, Lutterbaugh JD, Neumann A, Brattain MG, Chang J, Kim SJ, Kinzler KW, Vogelstein B, Willson JK and Markowitz S. (1999). *Cancer Res.*, **59,** 320–324.

Inman GJ and Allday MJ. (2000). *J. Gen. Vir.*, **81,** 1567–1578.

Ionov Y, Peinado MA, Malkhosyan S, Shibata D and Perucho M. (1993). *Nature*, **363,** 558–561.

Lu SL, Kawabata M, Imamura T, Akiyama Y, Nomizu T, Miyazono K and Yuasa Y. (1998). *Nat. Genet.*, **19,** 17–18.

Lu SL, Zhang WC, Akiyama Y, Nomizu T and Yuasa Y. (1996). *Cancer Res.*, **56,** 4595–4598.

Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW and Vogelstein B. (1995). *Science*, **268,** 1336–1338.

Mironov N, Jansen LA, Zhu WB, Aguelon AM, Reguer G and Yamasaki H. (1999). *Carcinogenesis*, **20,** 2189–2192.

Nicolaides NC, Littman SJ, Modrich P, Kinzler KW and Vogelstein B. (1998). *Mol. Cell Biol.*, **18,** 1635–1641.

Ouyang H, Shiwaku HO, Hagiwara H, Miura K, Abe T, Kato Y, Ohtani H, Shiiba K, Souza RF, Meltzer SJ and Horii A. (1997). *Cancer Res.*, **57,** 1851–1854.

Papadopoulos N and Lindblom A. (1997). *Hum. Mutat.*, **10,** 89–99.

Parsons R, Myeroff LL, Liu B, Willson JK, Markowitz SD, Kinzler KW and Vogelstein B. (1995a). *Cancer Res.*, **55,** 5548–5550.

Parsons R, Li GM, Longley M, Modrich P, Liu B, Berk T, Hamilton SR, Kinzler KW and Vogelstein B. (1995b). *Science*, **268,** 738–740.

Pedroni M, Sala E, Scarselli A, Borghi F, Menigatti M, Benatti P, Percesepe A, Rossi G, Foroni M, Losi L, Di Gregorio C, De Pol A, Nascimbeni R, Di Betta E, Salerni B, de Leon MP and Roncucci L. (2001). *Cancer Res.*, **61,** 896–899.

Rampino N, Yamamoto H, Ionov Y, Li Y, Sawai H, Reed JC and Perucho M. (1997). *Science*, **275,** 967–969.

Shibata D, Peinado MA, Ionov Y, Malkhosyan S and Perucho M. (1994). *Nat. Genet.*, **6,** 273–281.

Takenoshita S, Tani M, Nagashima M, Hagiwara K, Bennett WP, Yokota J and Harris CC. (1997). *Oncogene*, **14,** 1255–1258.

Vilkki S, Tsao JL, Loukola A, Poyhonen M, Vierimaa O, Herva R, Aaltonen LA and Shibata D. (2001). *Cancer Res.*, **61,** 4541–4544.

Wang CY, Eshleman JR, Willson JK and Markowitz S. (1995). *Cancer Res.*, **55,** 5101–5105.

Wheeler JM, Beck NE, Kim HC, Tomlinson IP, Mortensen NJ and Bodmer WF. (1999). *Proc. Natl. Acad. Sci. USA.*, **96,** 10296–10301.

Zhang L, Yu J, Willson JK, Markowitz SD, Kinzler KW and Vogelstein B. (2001). *Cancer Res.*, **61,** 3801–3805.